An introduction to the ORE interoperability framework

<http://www.openarchives.org/ore/toc>

Herbert Van de Sompel - herbertv@lanl.gov

Digital Library Research & Prototyping Team
Research Library
Los Alamos National Laboratory, USA

OAI-ORE is funded
by the Andrew W. Mellon Foundation, the National Science Foundation, JISC, and Microsoft

The ORE Editors are: Carl Lagoze (Cornell U.), Herbert Van de Sompel (LANL), Pete Johnston (Eduserv Found.), Michael Nelson (Old Dominion University), Robert Sanderson (U. of Liverpool), Simeon Warner (Cornell U.)

OAI Object Reuse and Exchange: Support

- The Andrew W. Mellon Foundation
- The Coalition for Networked Information
- Joint Information Systems Committee
- Microsoft Corporation
- The National Science Foundation

OAI Object Reuse and Exchange: Technical Experts

ORE Technical Committee

Chris Bizer	Freie Universität Berlin
Les Carr	University of Southampton
Tim DiLauro	Johns Hopkins University
Leigh Dodds	Ingenta
David Fulker	UCAR
Tony Hammond	Nature Publishing Group
Pete Johnston	Eduserv Foundation
Richard Jones	HP Labs
Carl Lagoze	Cornell University
Peter Murray	OhioLINK
Michael Nelson	Old Dominion University
Ray Plante	NCSA and National Virtual Observatory
Rob Sanderson	University of Liverpool
Herbert Van de Sompel	Los Alamos National Laboratory
Simeon Warner	Cornell University
Jeff Young	OCLC

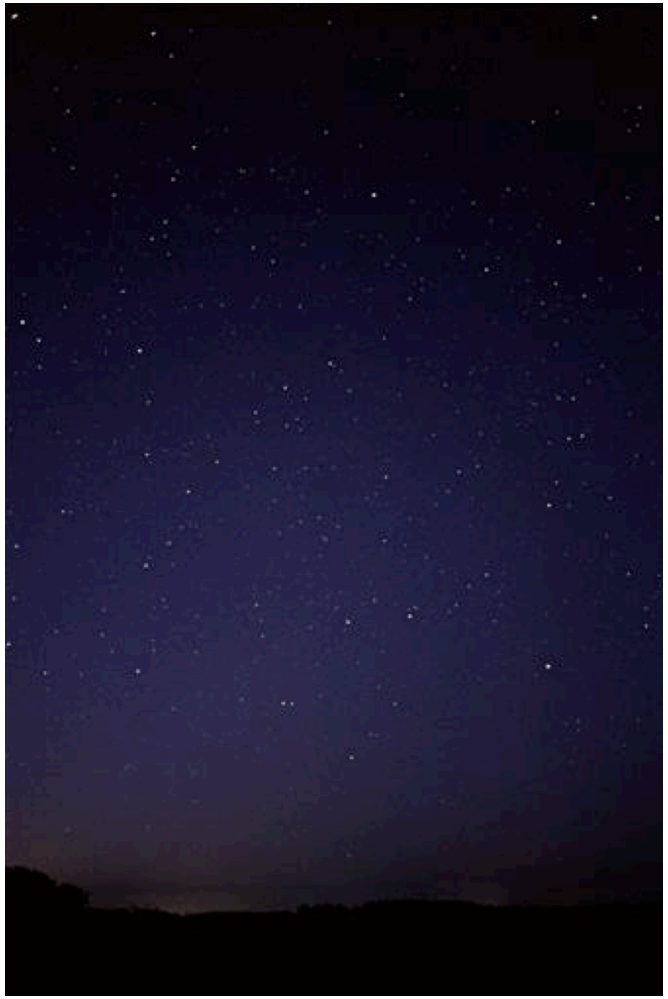
ORE Liaison Group

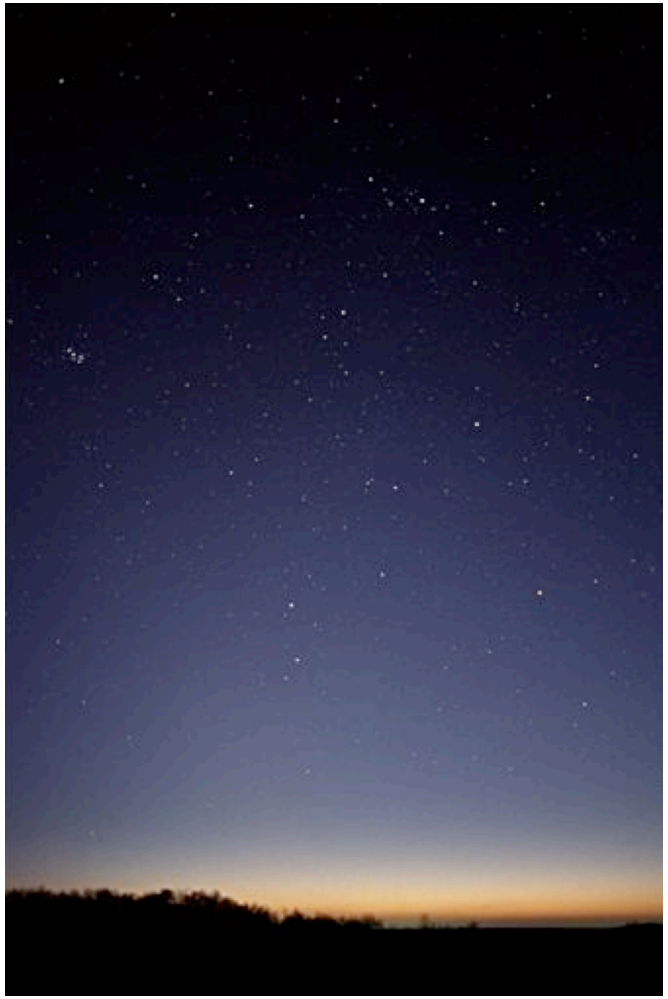
Leonardo Candela	Consiglio Nazionale delle Ricerche - DRIVER
Tim Cole	University of Illinois Urbana-Champaign - Aquifer
Julie Allinson	JISC
Jane Hunter	University of Queensland - DEST
Savas Parastatidis	Microsoft Corporation
Sandy Payette	Fedora Commons
Thomas Place	University of Tilburg - DARE
Andy Powell	Eduserv Foundation - DCMI
Robert Tansley	Google, Inc. - DSpace

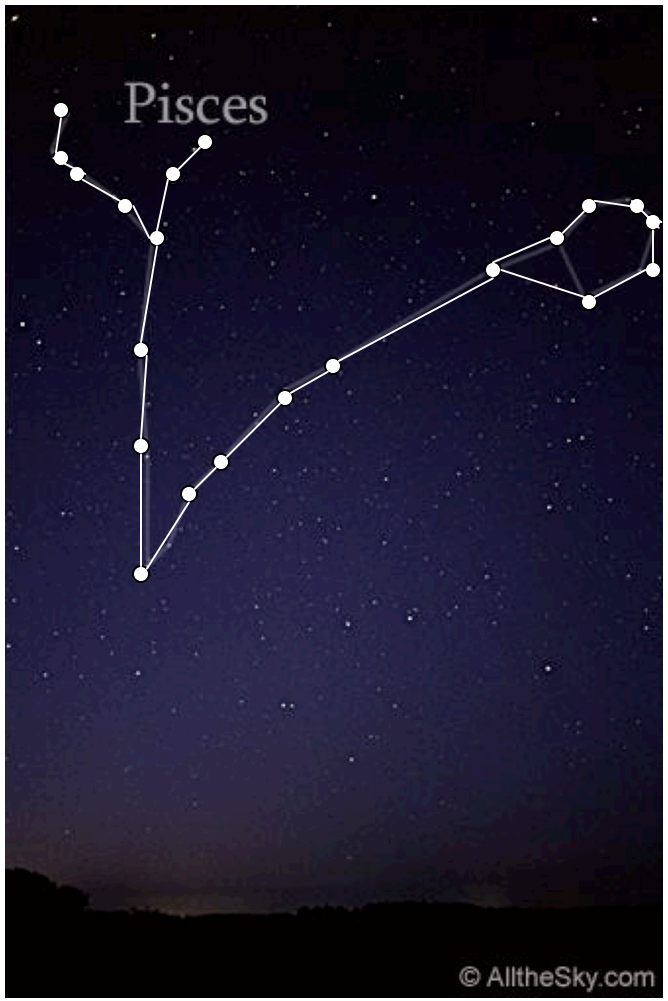
OAI Object Reuse and Exchange

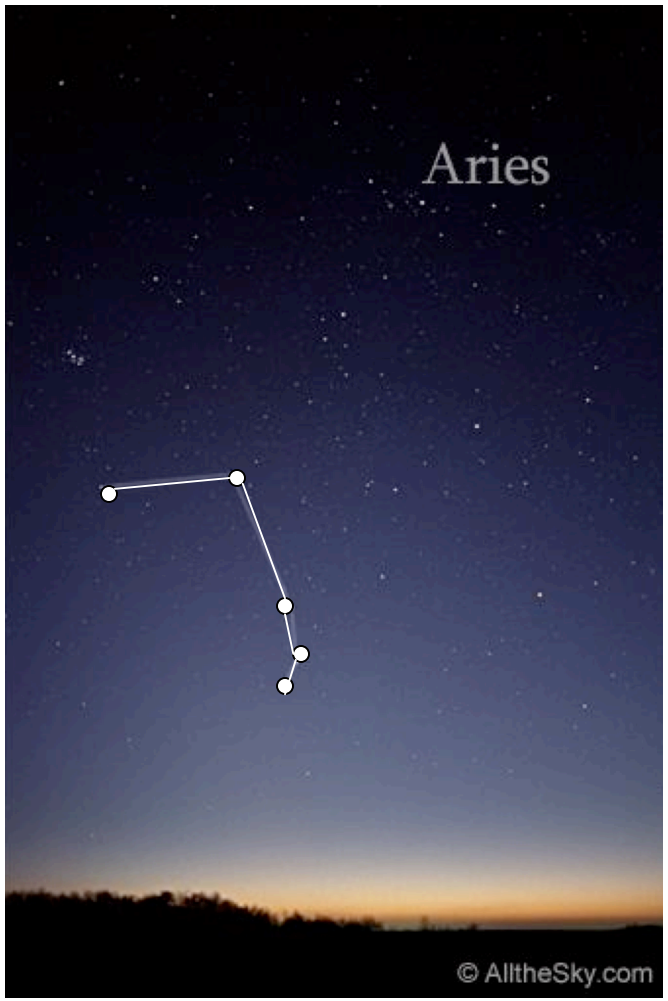
Subject: **Aggregations** of Web resources

Approach: Publish **Resource Maps** to the Web that Instantiate, Describe, and Identify Aggregations



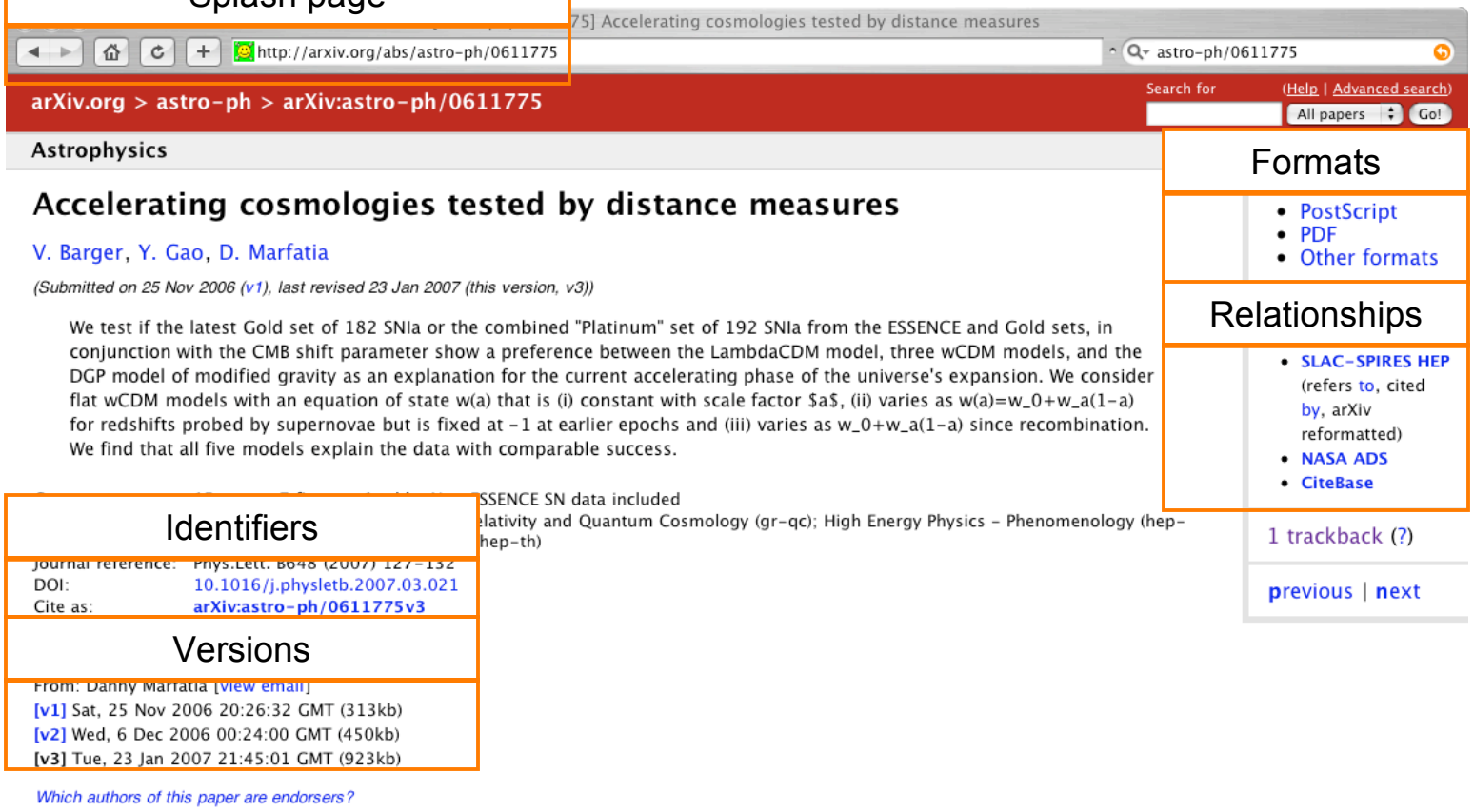






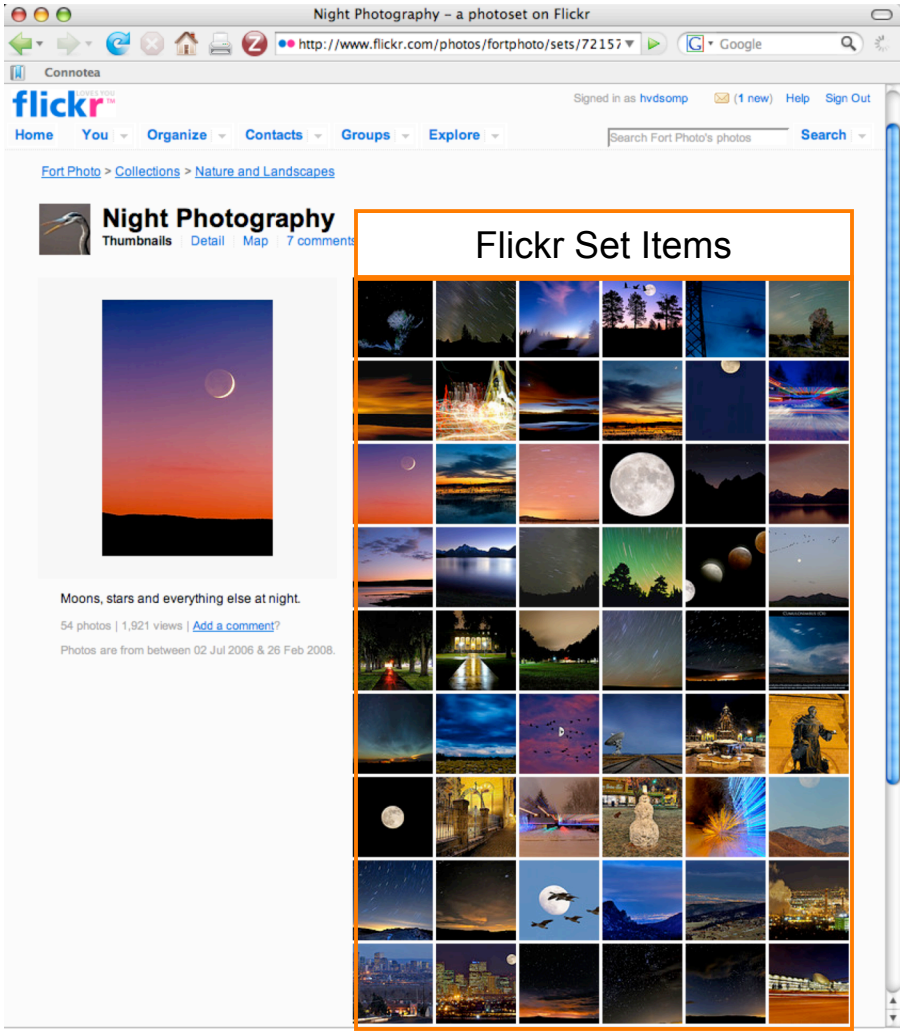
Instantiate, Describe, and Identify Aggregations

Aggregations!



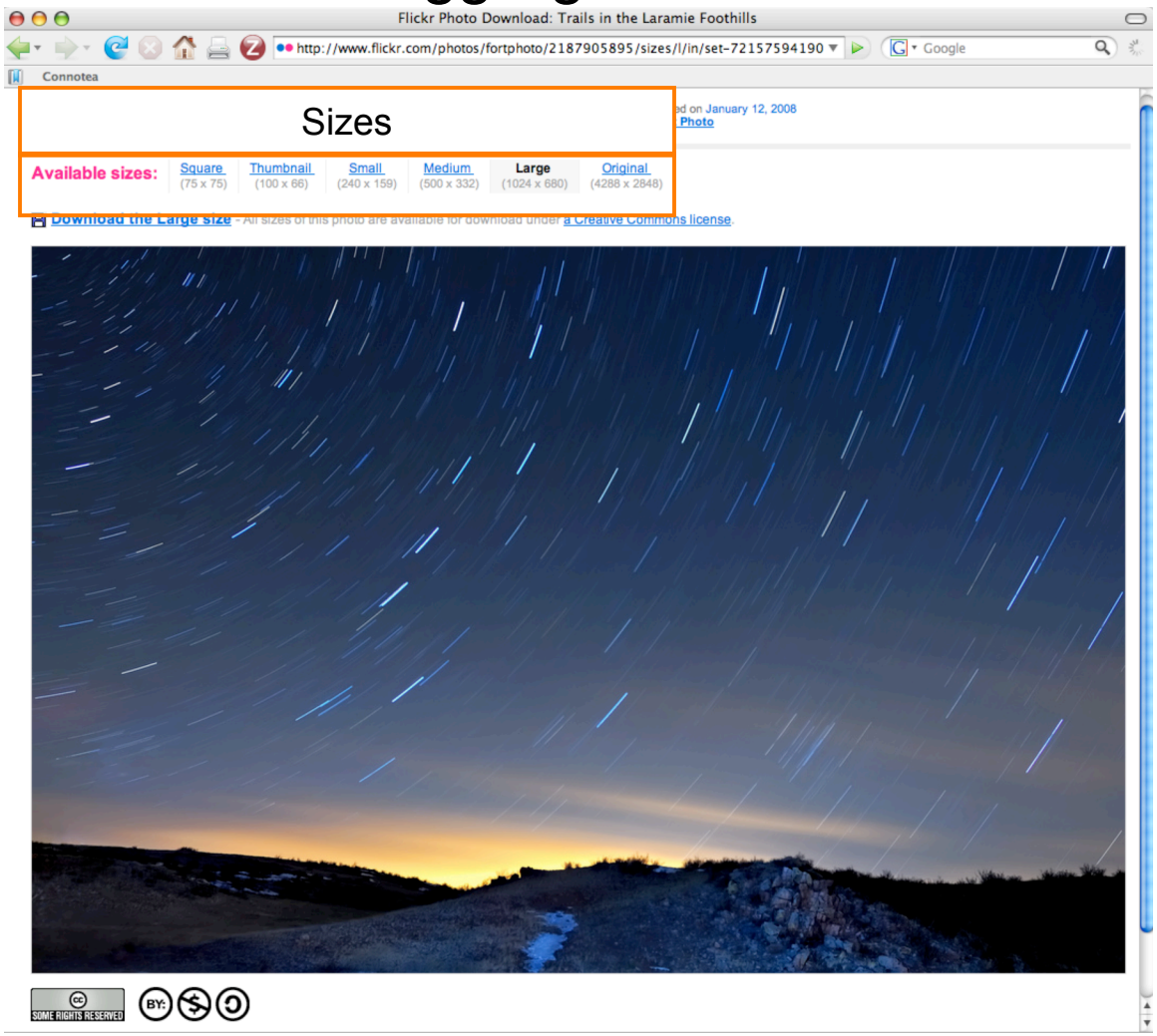
http://arxiv.org/abs/astro-ph/0611775

Aggregations!!



http://www.flickr.com/photos/fortphoto/sets/72157594190371016/

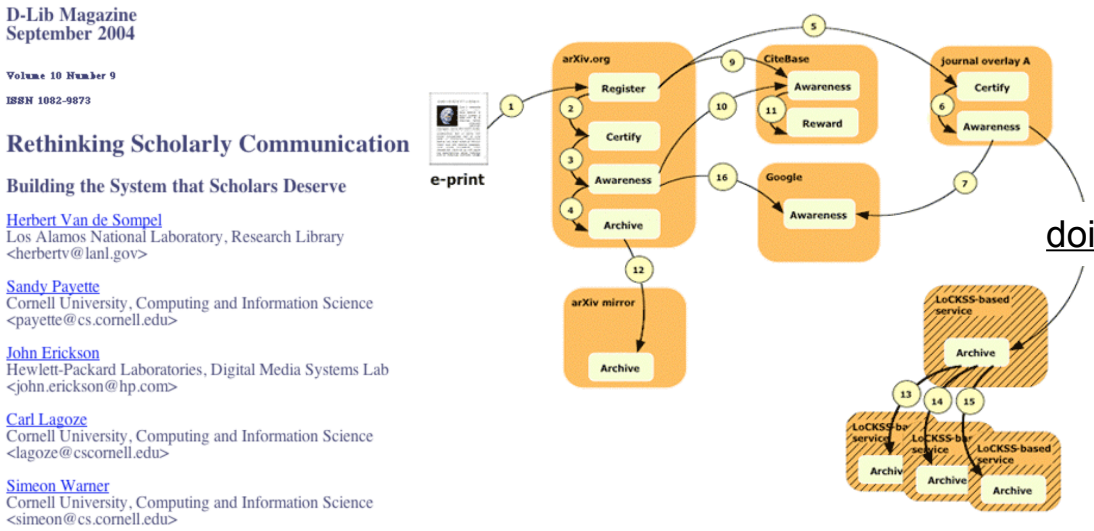
Aggregations!!!



http://www.flickr.com/photos/fortphoto/sets/72157594190371016/

OAI Object Reuse and Exchange: Original Vision

- Scholarly communication as a global, cross-repository workflow.
  - Leverage the intrinsic value of the materials that become available in distributed repositories.
  - Value chains across repositories and applications with repository materials as their subject.
  - Make repositories **active nodes in a global environment**, not passive local nodes.
  - Life for those materials **starts** in repositories; it does not end there.
  - Materials from repositories must be **reusable in different contexts**.



D-Lib Magazine
September 2004

Volume 10 Number 9

ISSN 1082-9873

Rethinking Scholarly Communication

Building the System that Scholars Deserve

Herbert Van de Sompel
Los Alamos National Laboratory, Research Library
<herbertv@lanl.gov>

Sandy Payette
Cornell University, Computing and Information Science
<payette@cs.cornell.edu>

John Erickson
Hewlett-Packard Laboratories, Digital Media Systems Lab
<john.erickson@hp.com>

Carl Lagoze
Cornell University, Computing and Information Science
<lagoze@cscornell.edu>

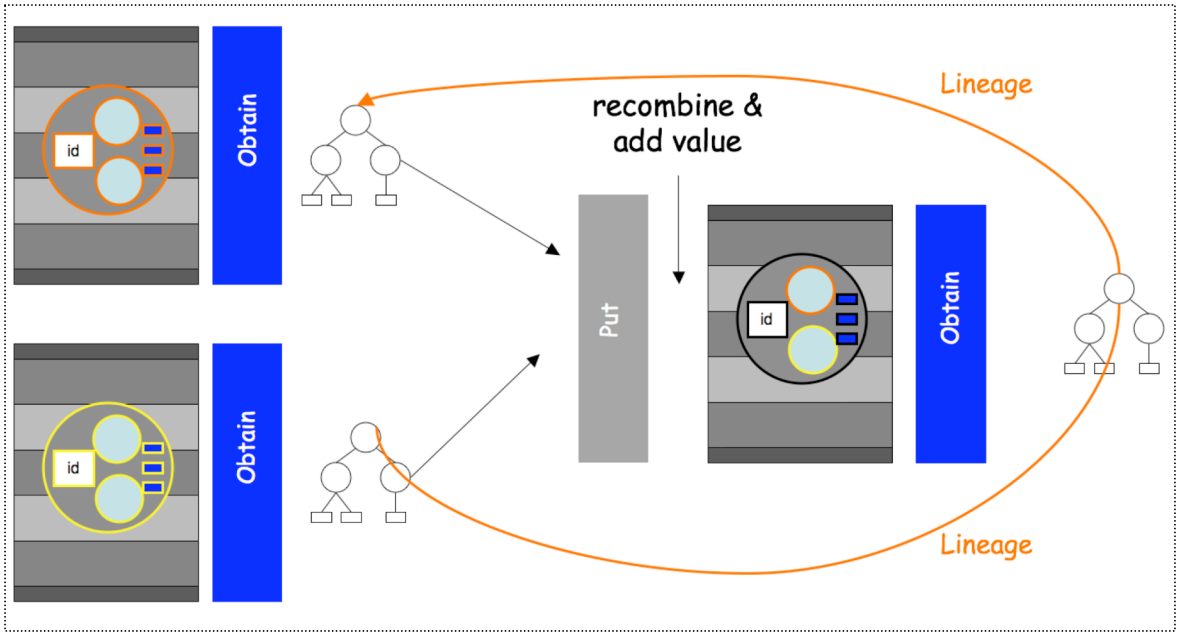
Simeon Warner
Cornell University, Computing and Information Science
<simeon@cs.cornell.edu>

doi:10.1045/september2004-vandesompel

Scholarly Communication as a Global Workflow

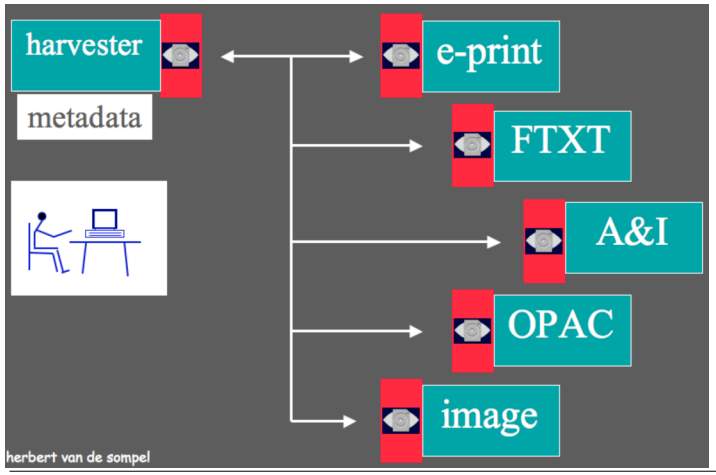
- That vision translated into fashionable contemporary terminology amounts to enabling Web 2.0 scholarly mash-ups
  - augmented with the ability to deal with compound information objects
  - augmented with some capability to trace the lineage of resources used in the workflow
  - augmented with some capability to record the nature of the workflow (reproduce-ability)

A slide from the archive (2005)

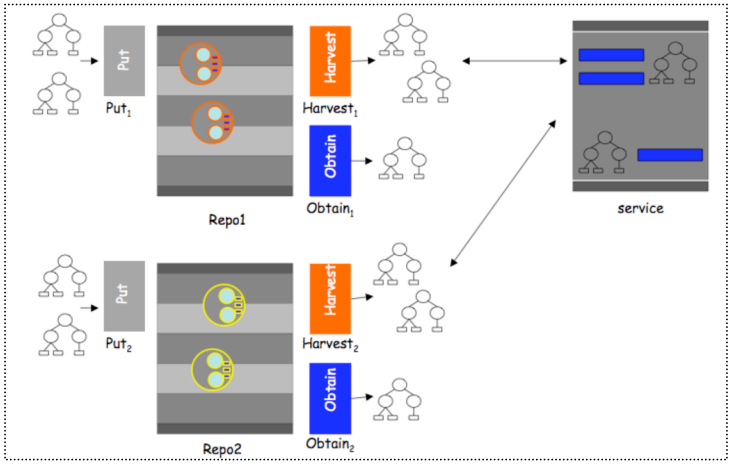


Object Reuse and Exchange: A Resource-Centric Approach

- Prior efforts had the *repository* and *metadata records* as the center of the interoperability thinking:
  - Including OAI-PMH
  - Including initial OAI-ORE thinking cf. “Augmenting Interoperability across Scholarly Repositories”
  - Unclear what the metadata records were about …
- This approach does not vibe well with the Web:
  - The Web Architecture knows resources and URIs
  - Requires special treatment by applications that dominate the Web.

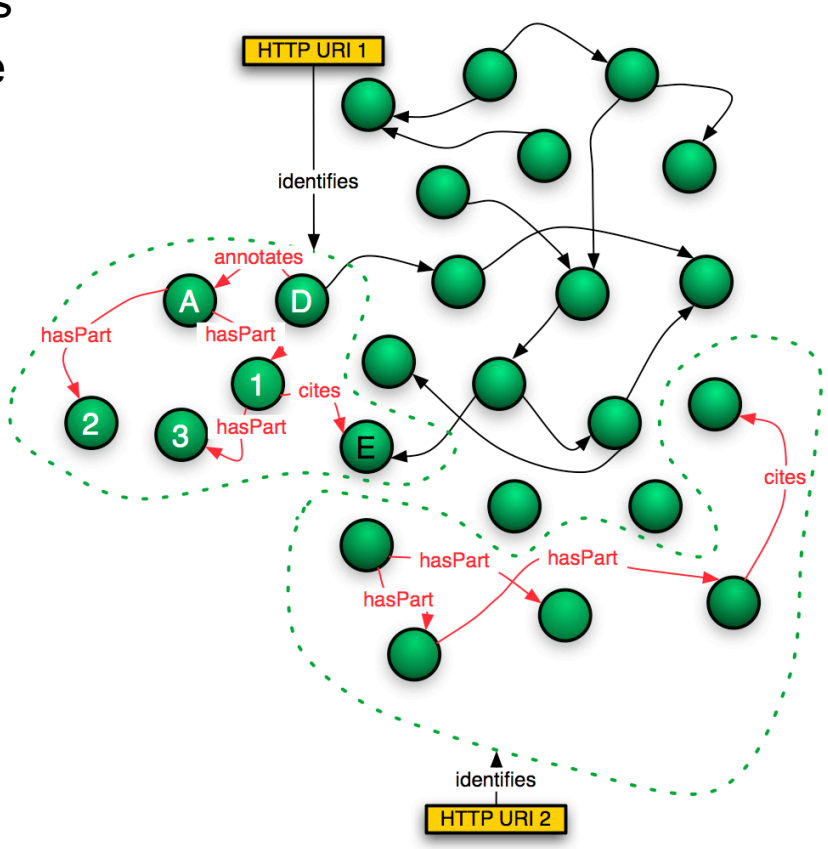


Keep dreaming!



Object Reuse and Exchange: A Resource-Centric Approach

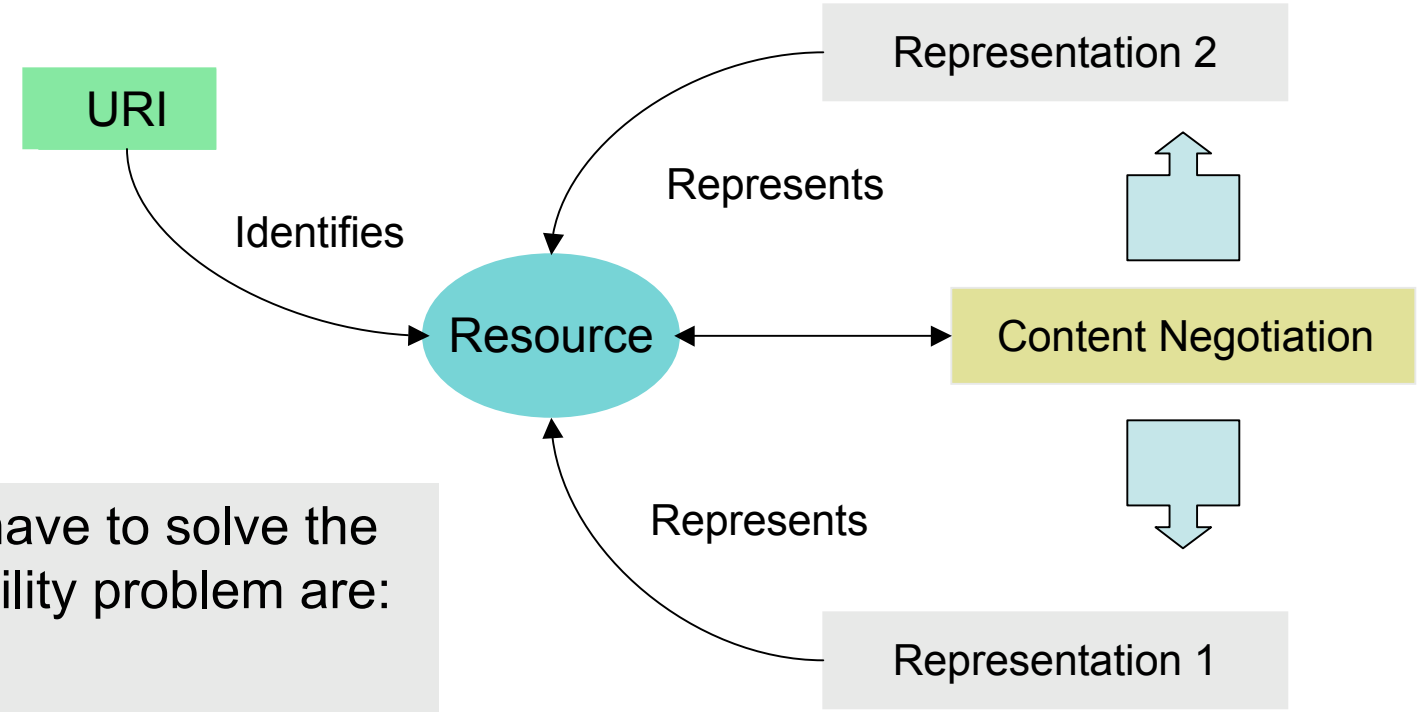
- Fundamental shift in the chosen approach towards interoperability
- The Web Architecture as the platform for interoperability
- Resources, URIs, and representations as the tools of the ORE interoperability trade
- De-facto integration with existing Web applications
- Potential of adoption by other communities
- Potential of tools created by other communities
- ….



Object Reuse and Exchange: Towards a Machine-Readable Web

- Major inspiration/foundation for ORE:
  - Web Architecture
    - <http://www.w3.org/TR/webarch/>
  - Semantic Web, RDF
    - <http://www.w3.org/TR/rdf-primer/>
  - Linked Data
    - <http://linkeddata.org/>
    - <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>
  - Cool URIs for the Semantic Web
    - <http://www.w3.org/TR/cooluris>
  - [Named Graphs
    - <http://www.w3.org/2004/03/trix/>]

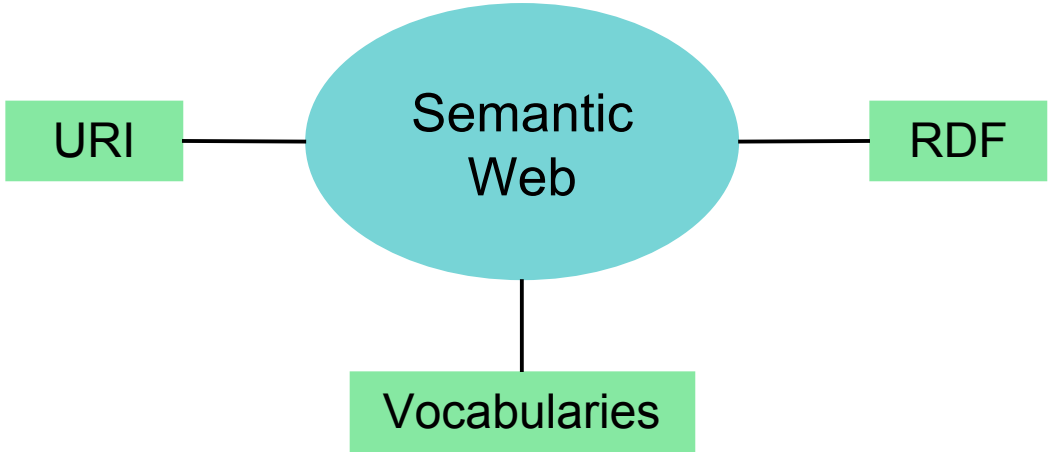
W3C Web Architecture



The tools we have to solve the interoperability problem are:

- Resource
- URI
- Representation

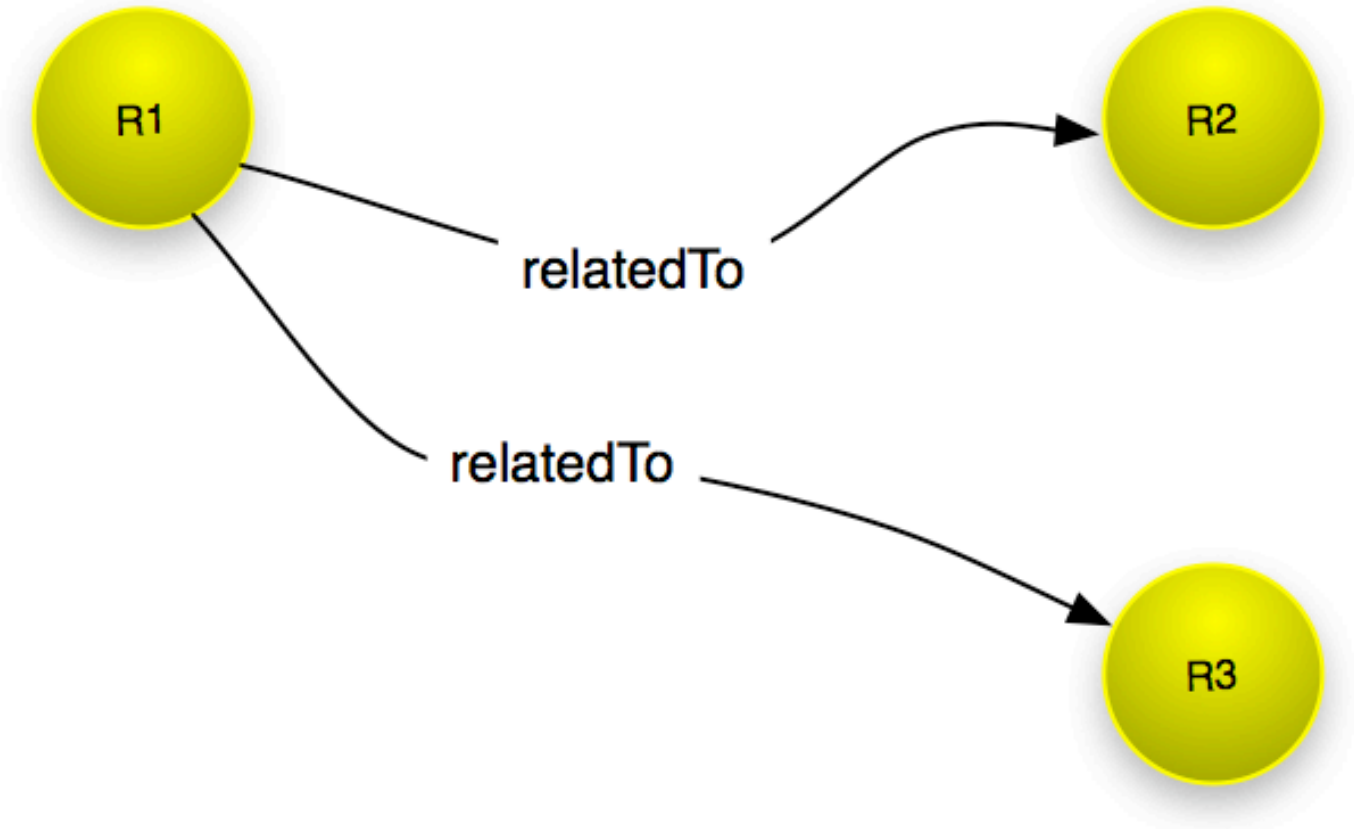
Semantic Web



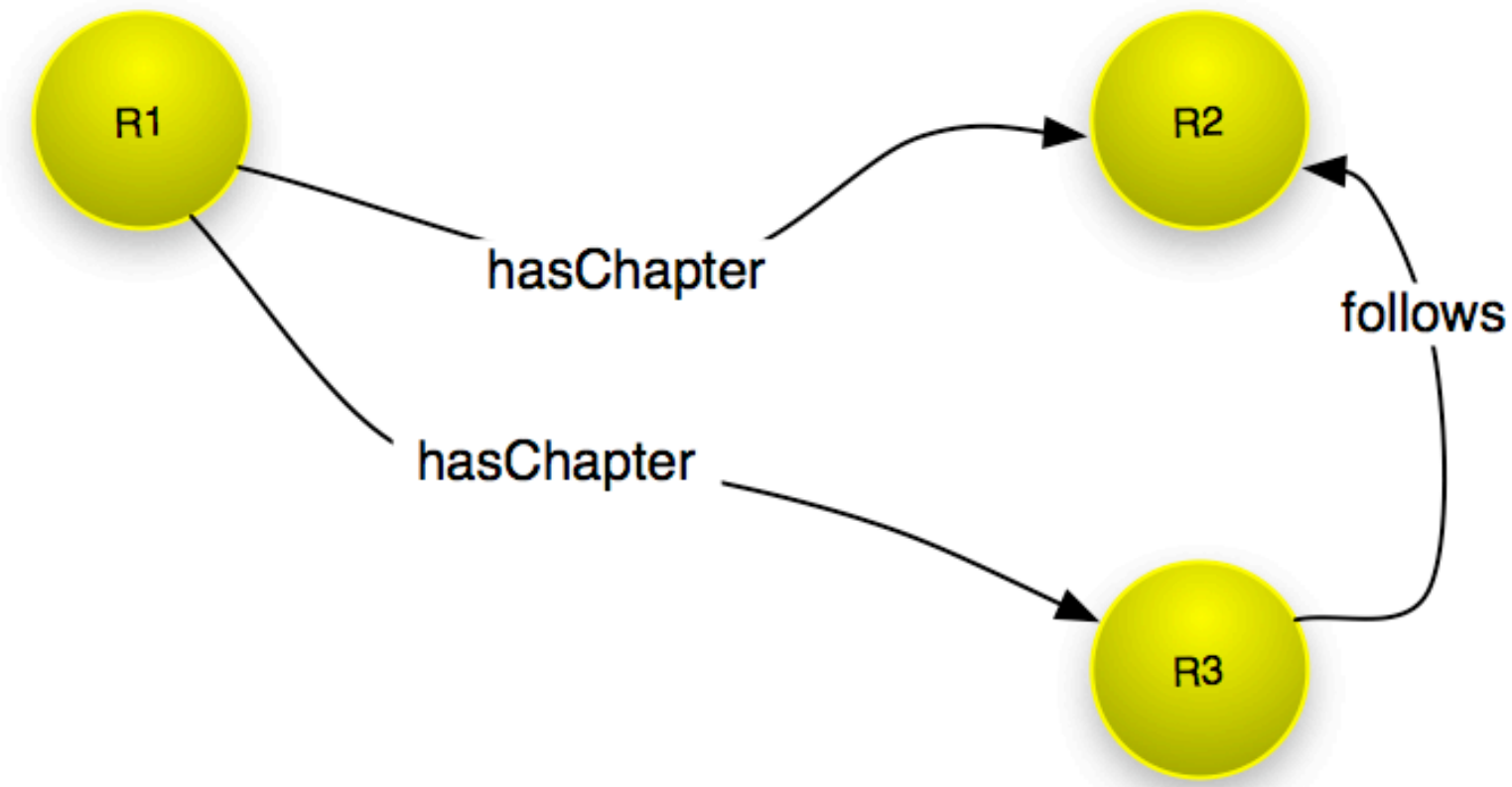
The tools we have to solve the interoperability problem are:

- URI
- RDF
- Vocabularies

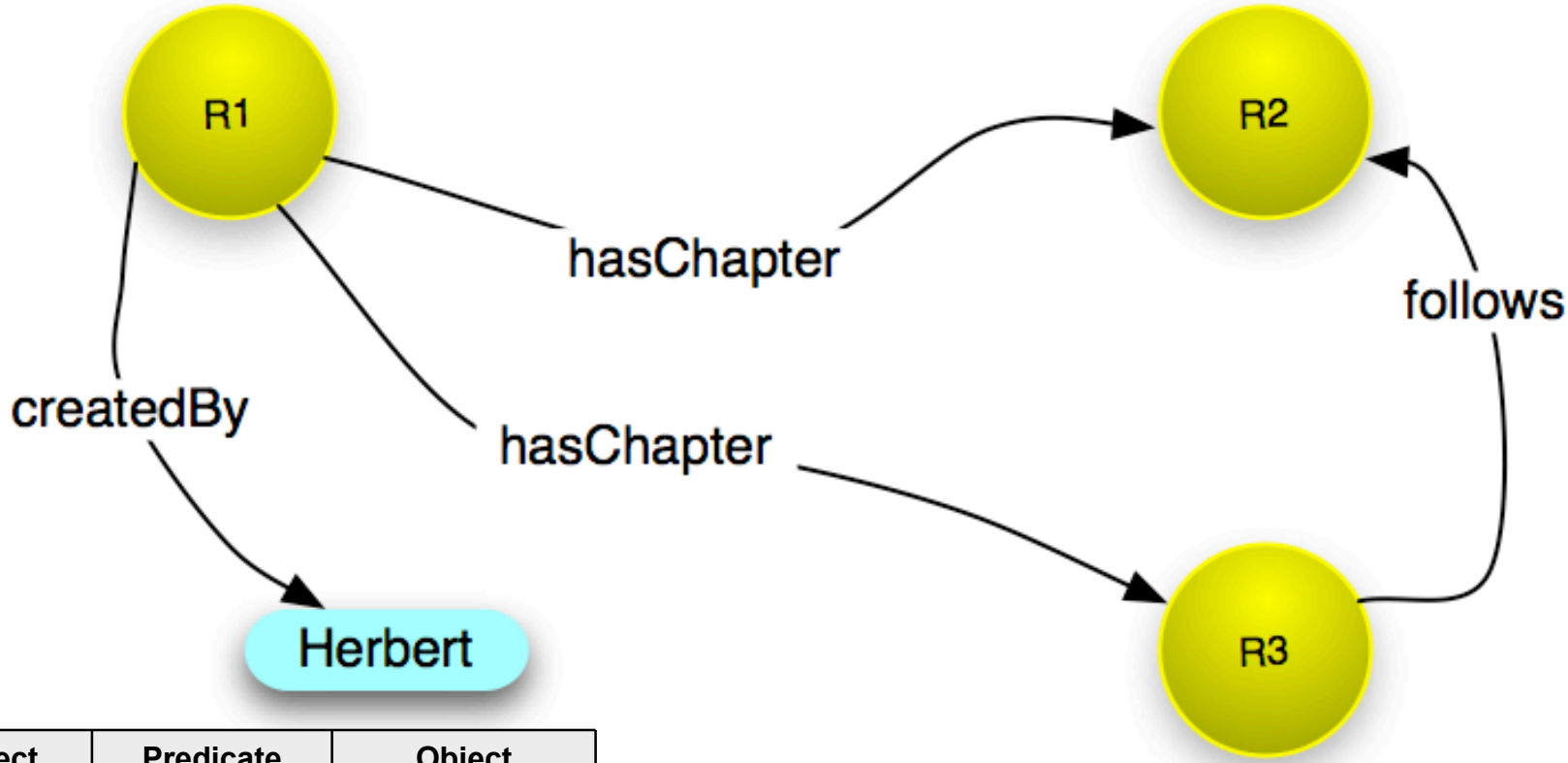
Resource Description Framework (RDF)



Resource Description Framework (RDF)



Resource Description Framework (RDF)



Subject	Predicate	Object
R1	hasChapter	R2
R1	hasChapter	R3
R3	follows	R2
R1	createdBy	“Herbert”

Triples

Linked Data: *RDF Lite over HTTP ;-)*

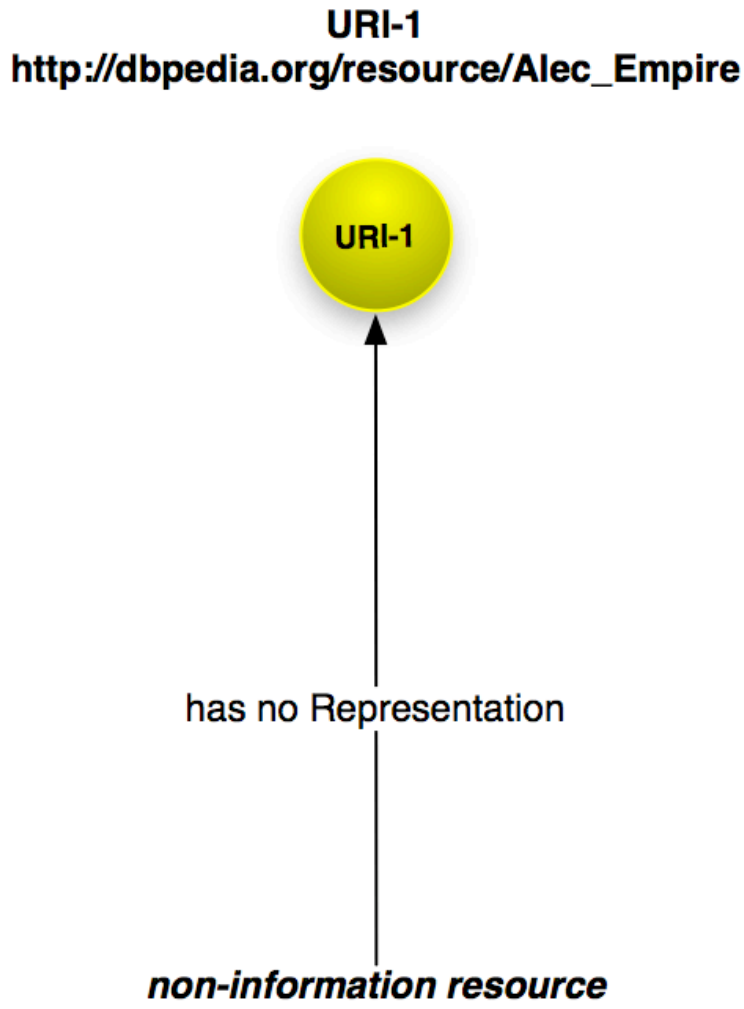
- Berners-Lee's Linked Data principles:
  1. Use URIs as names for things.
  2. Use HTTP URIs so that people can look up those names.
  3. When someone looks up a URI, provide useful information.
  4. Include links to other URIs. So that they can discover more things.

- Bizer's recommendations re which RDF features **NOT TO USE** in order to make it easier for clients to merge and query your data:
  - Blank nodes (not externally link-able, data merging issues)
  - Reification (unclear semantics, SPARQL issues) - use metadata about the information resource instead
  - Containers, Collections (SPARQL issues)

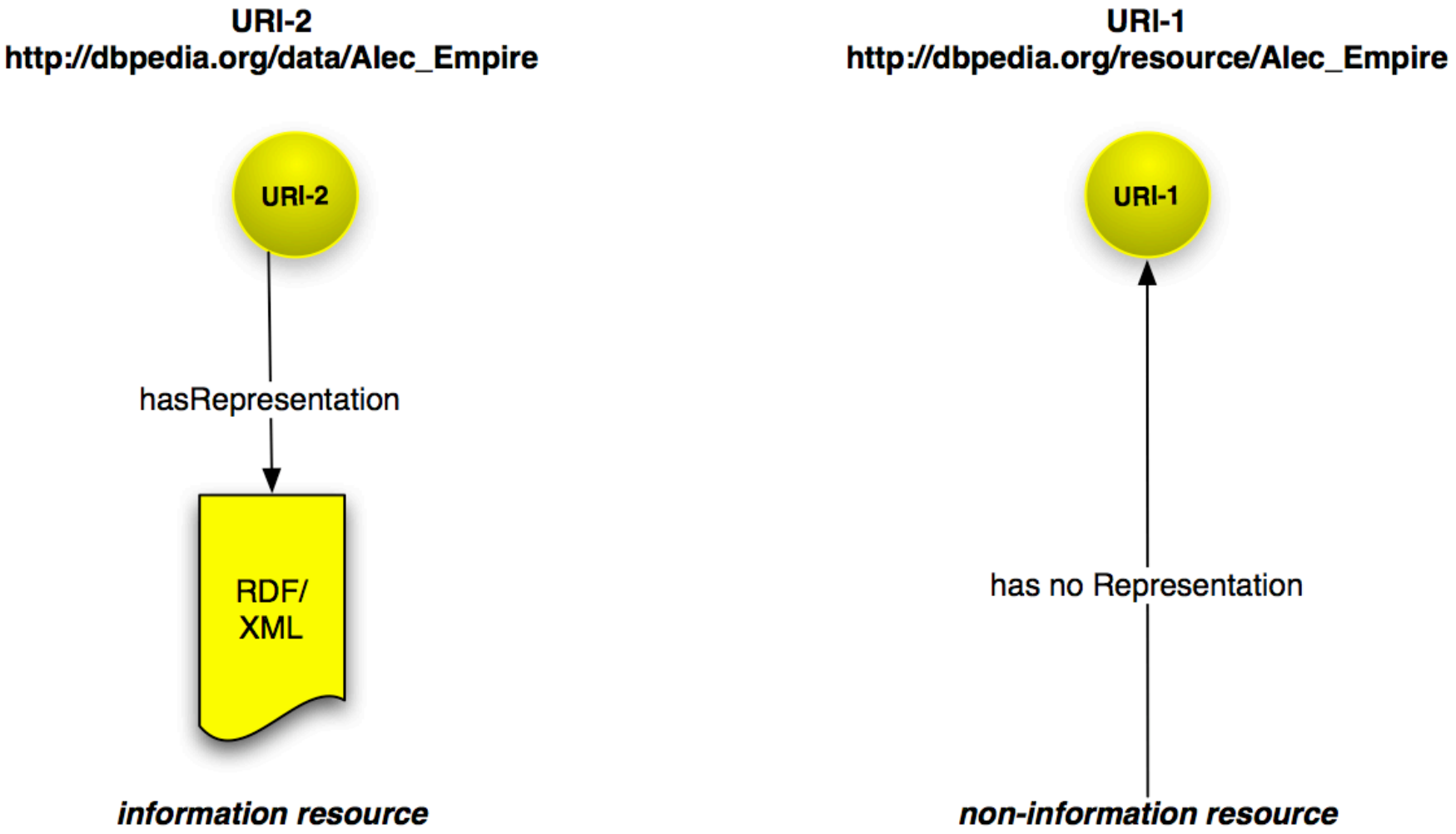

Linked Data: *RDF Lite over HTTP ;-)*

- Bizer's recommendations re the useful information to return when accessing the URI of a resource:
  - The description: all triples with the URI as subject
  - Backlinks: all triples with the URI as object (sometimes redundant but allows bidirectional traversal)
  - Related descriptions: triples about resources related to the resource
  - Metadata: information about the information that is returned, e.g. authorship, rights, publication datetime, etc.
  - Syntax: at least RDF/XML

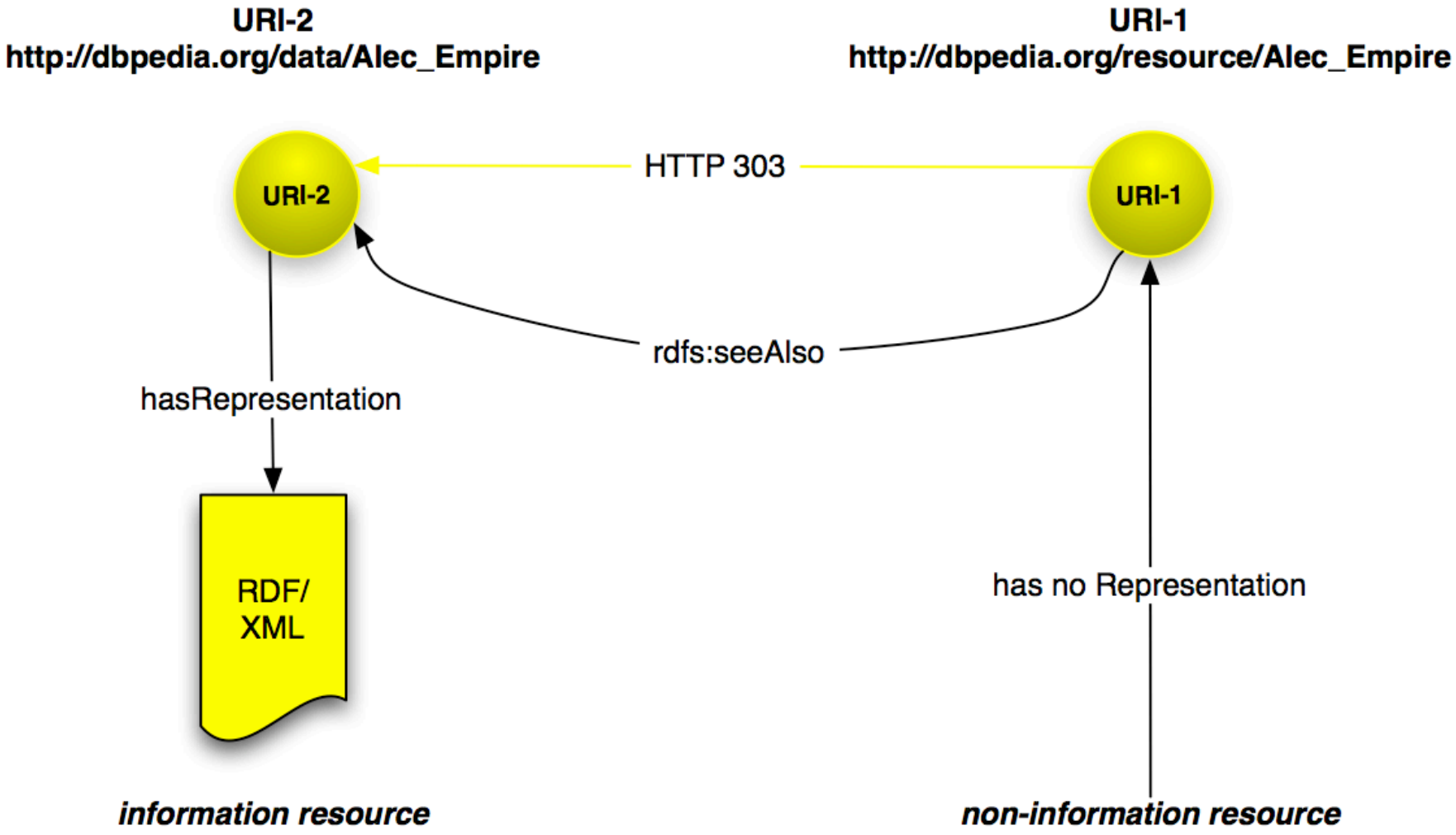

Cool URIs for the Semantic Web



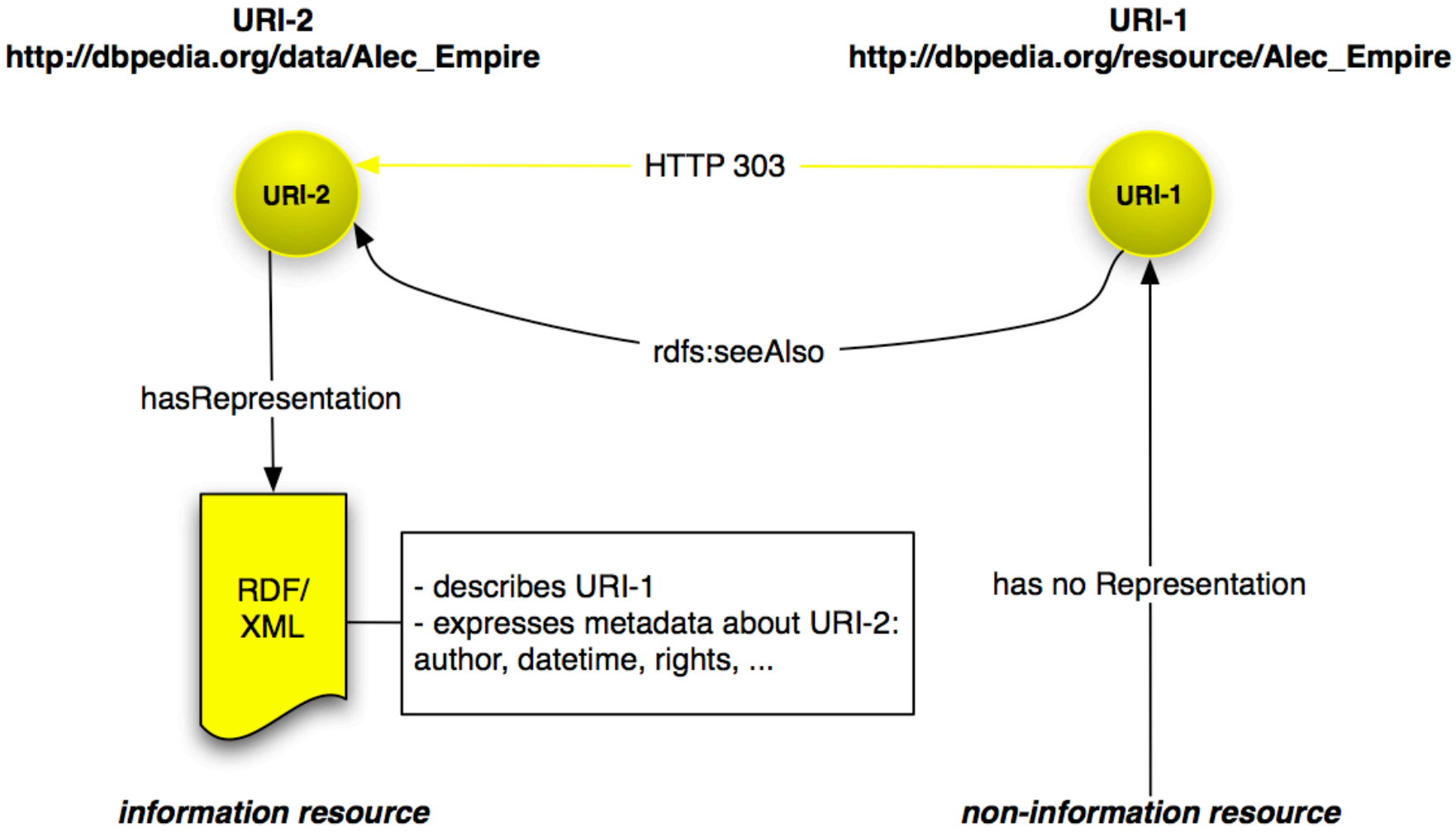
Cool URIs for the Semantic Web



Cool URIs for the Semantic Web



Cool URIs for the Semantic Web



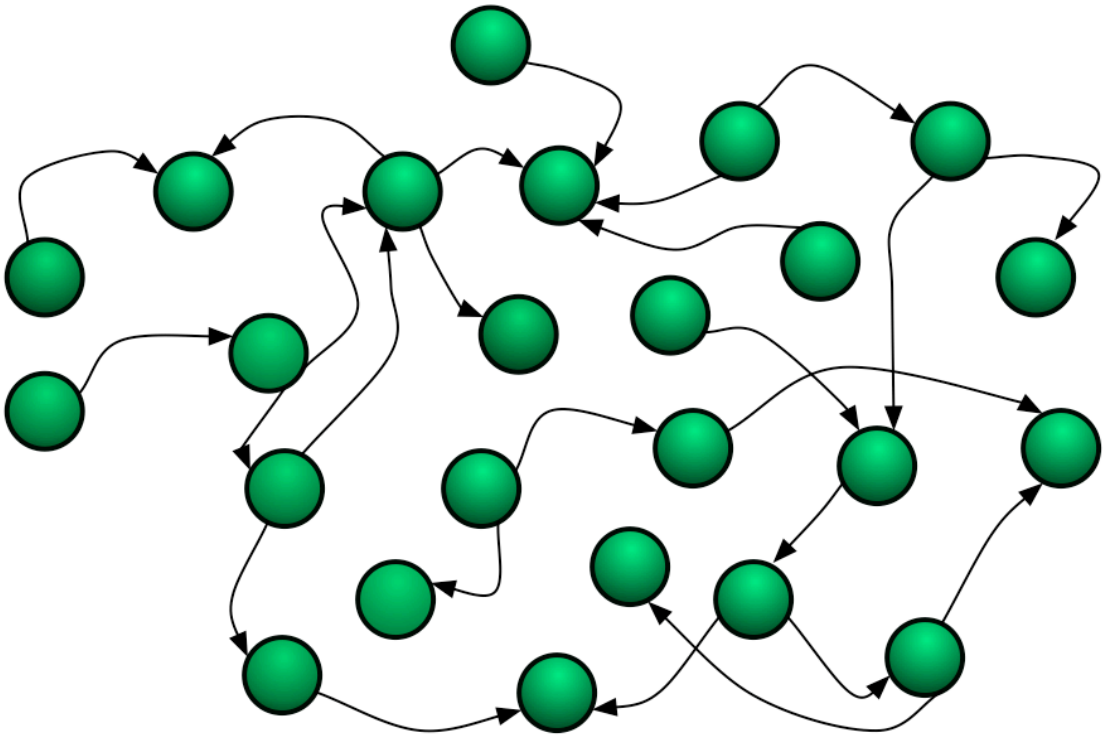
OAI Object Reuse and Exchange: The Approach

Subject: **Aggregations** of Web resources

Approach: Publish **Resource Maps** to the Web that Instantiate, Describe, and Identify Aggregations

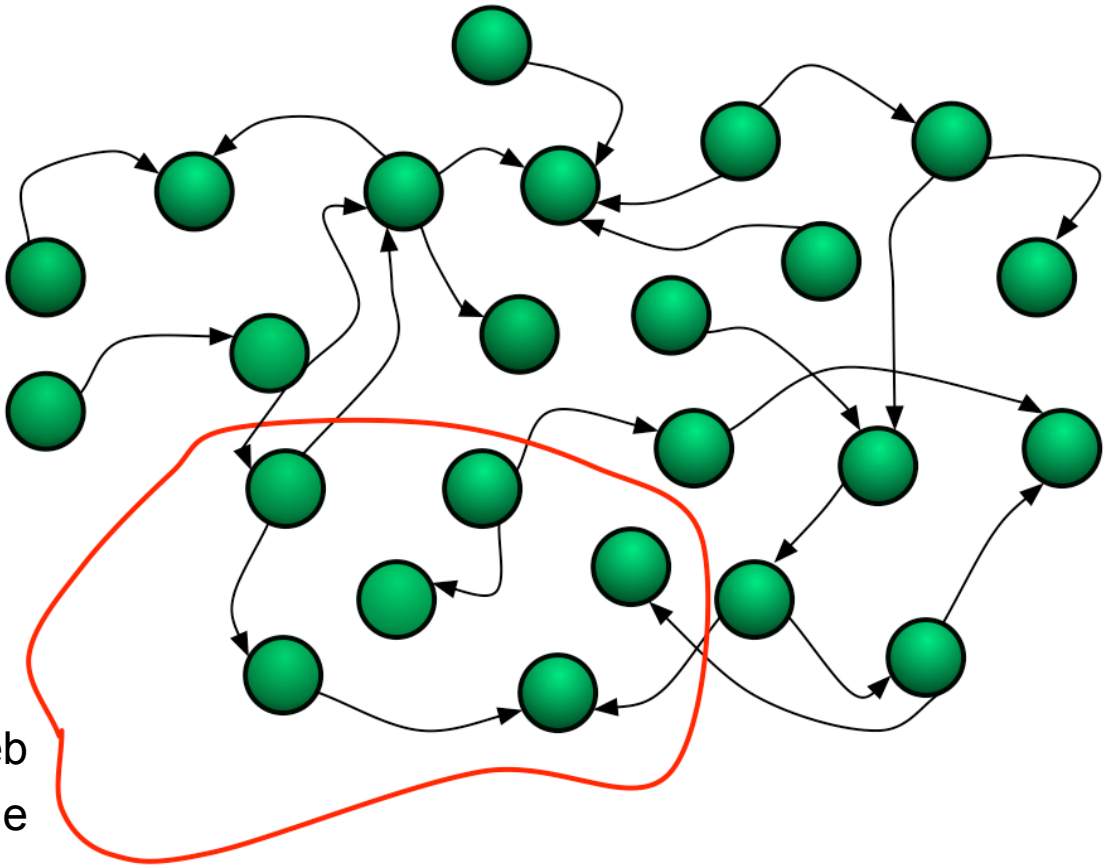
Aggregations: *non-information resources*
Resource Maps: *information resources*

The Web



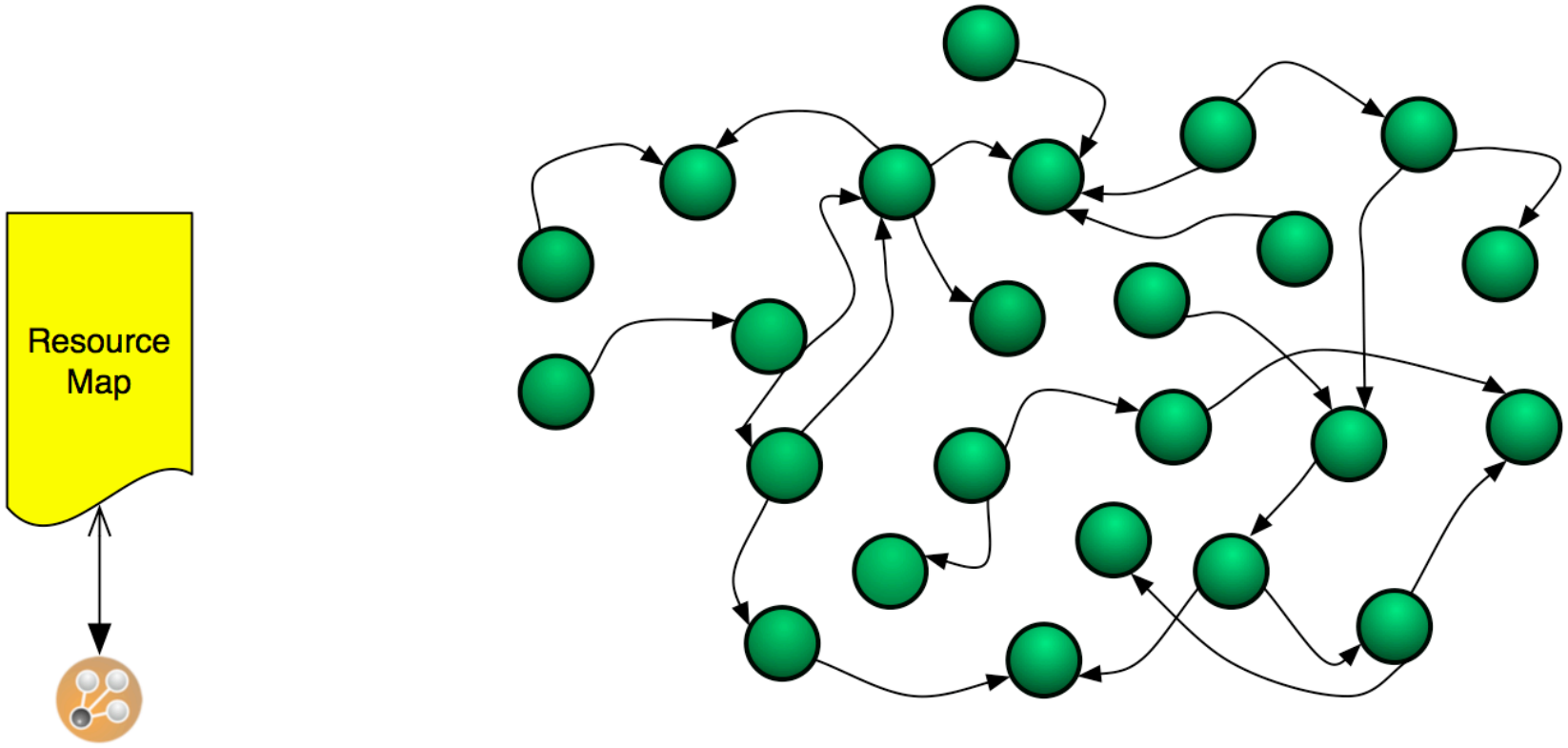
An Aggregation and the Web

- Resources of an Aggregation are distinct URI-identified Web resources

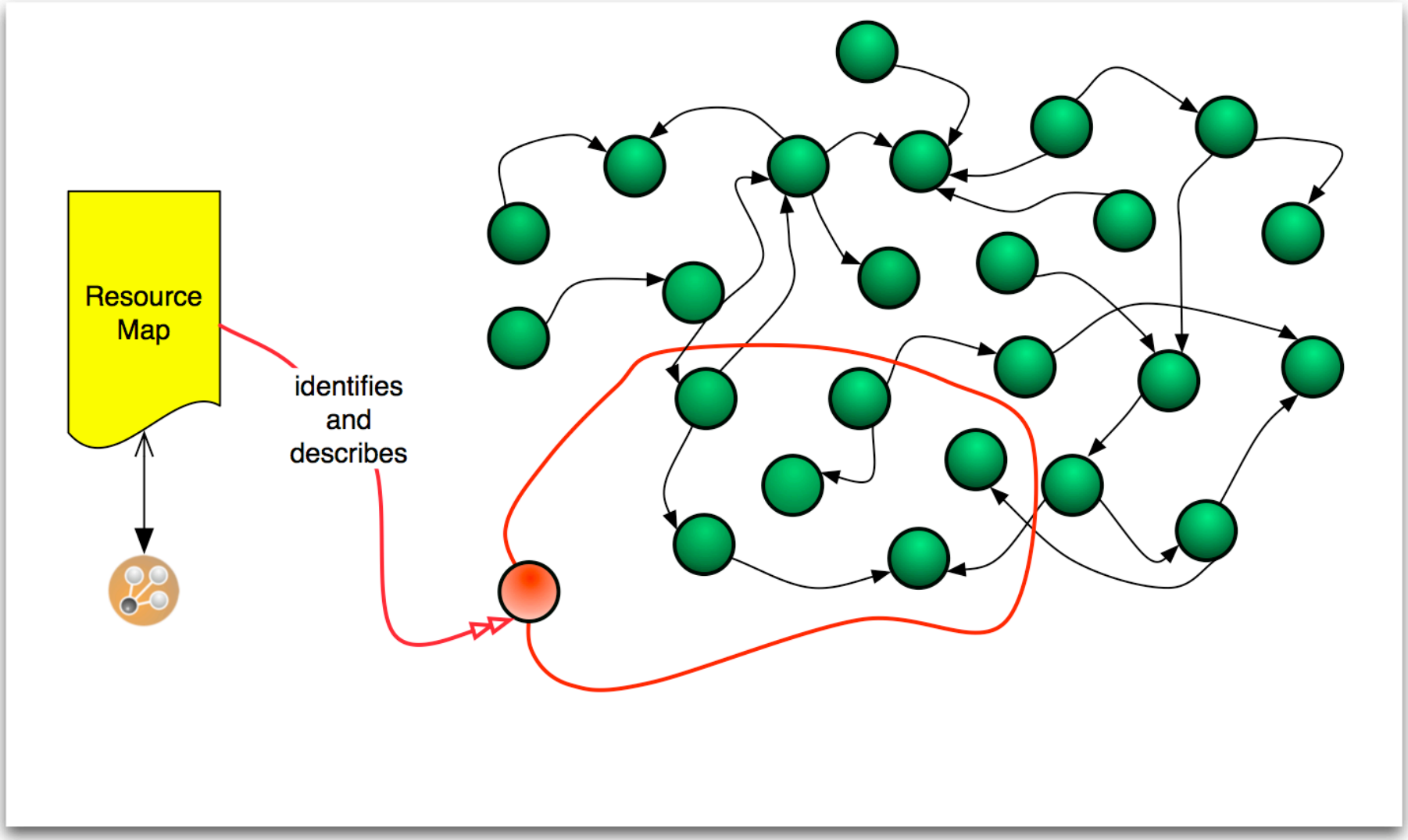


- Missing are:
  - The boundary that delineates the Aggregation in the Web
  - An identity (URI) for the Aggregation

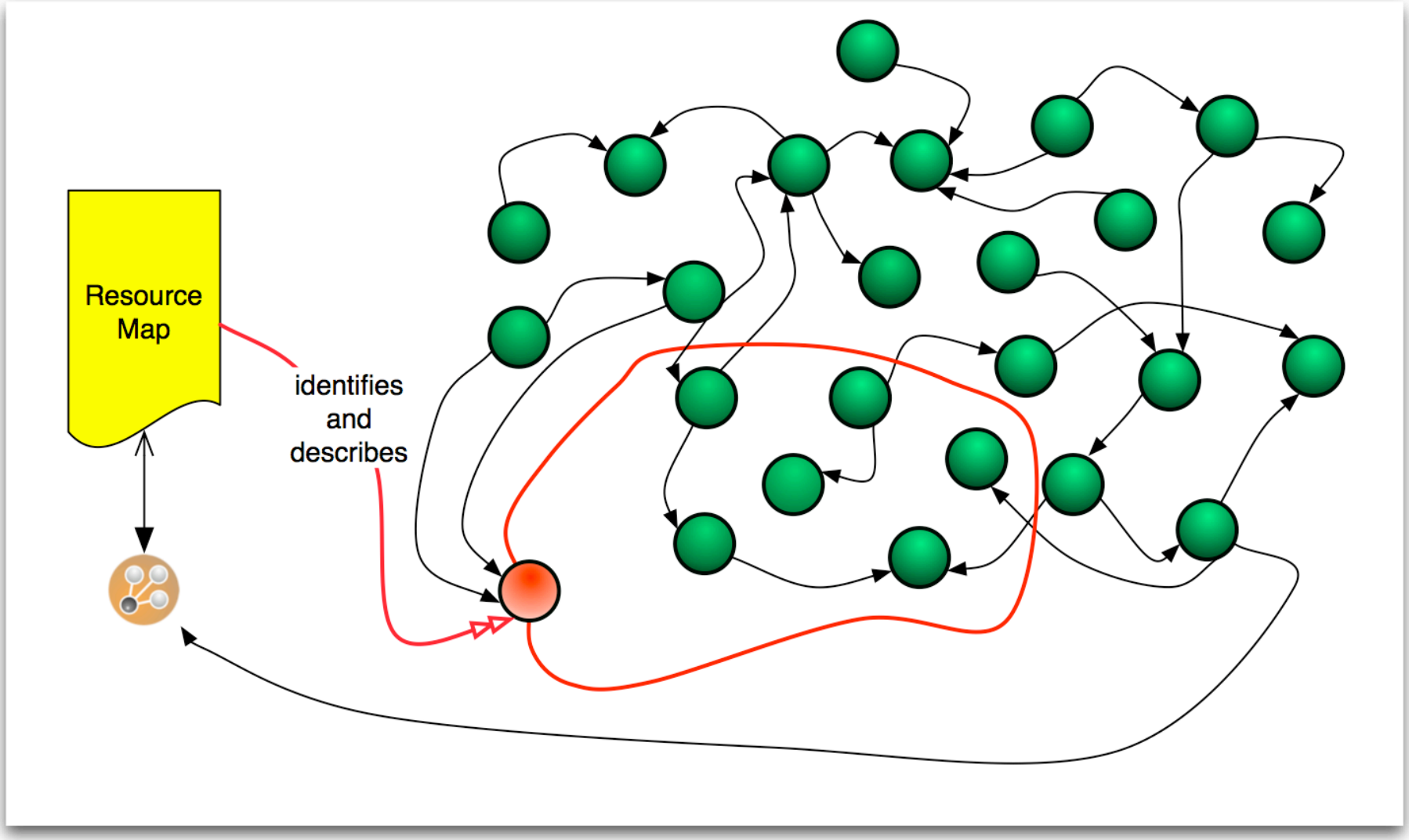
Publish a Resource Map to the Web



The Resource Map Identifies and Describes the Aggregation



The Resource Map and the Aggregation integrate into the Web



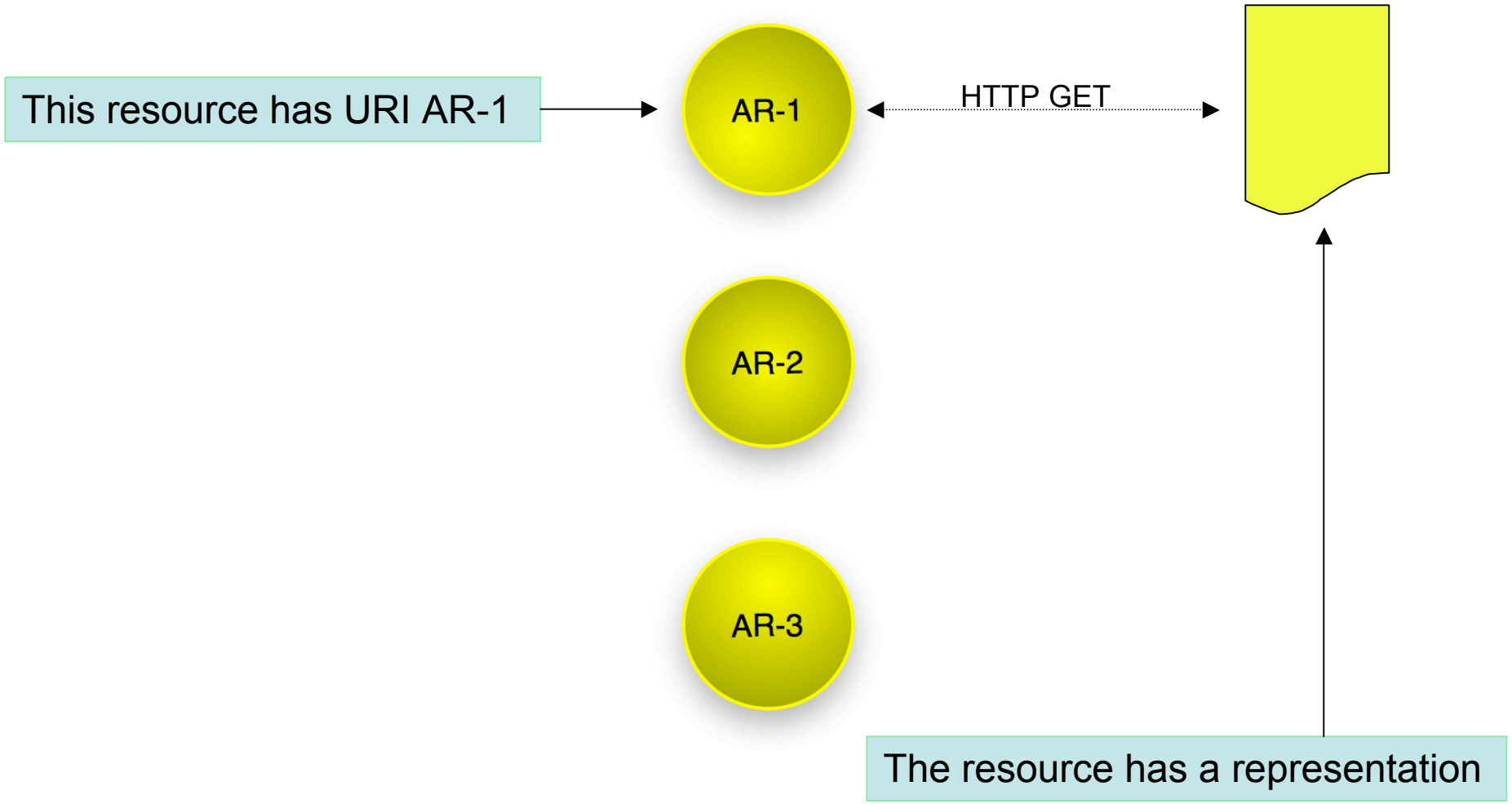
OAI Object Reuse and Exchange: The Basics

Aggregation
Aggregated Resources
`ore:aggregates`

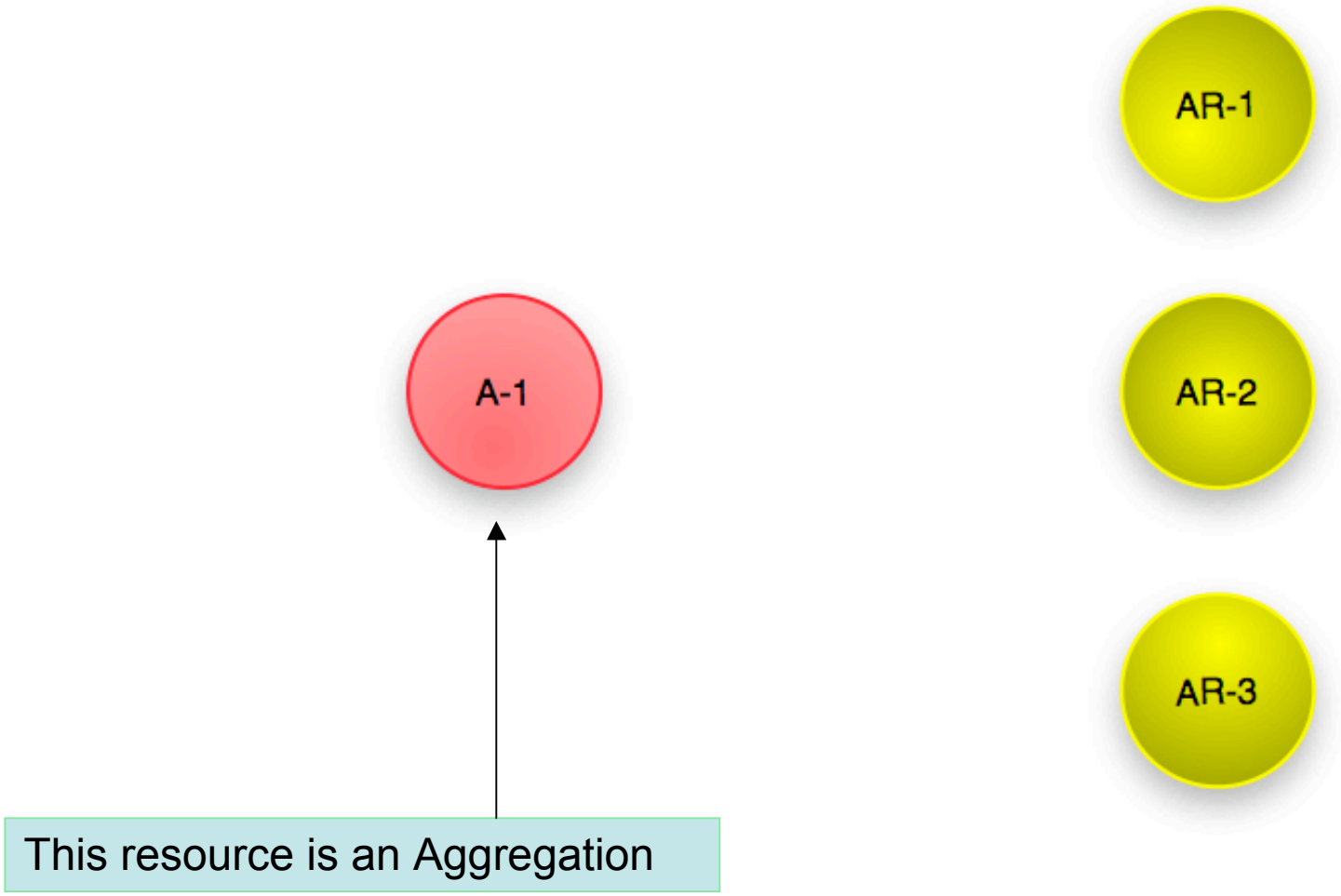
Resource Map
`ore:describes`

Relationships and Types

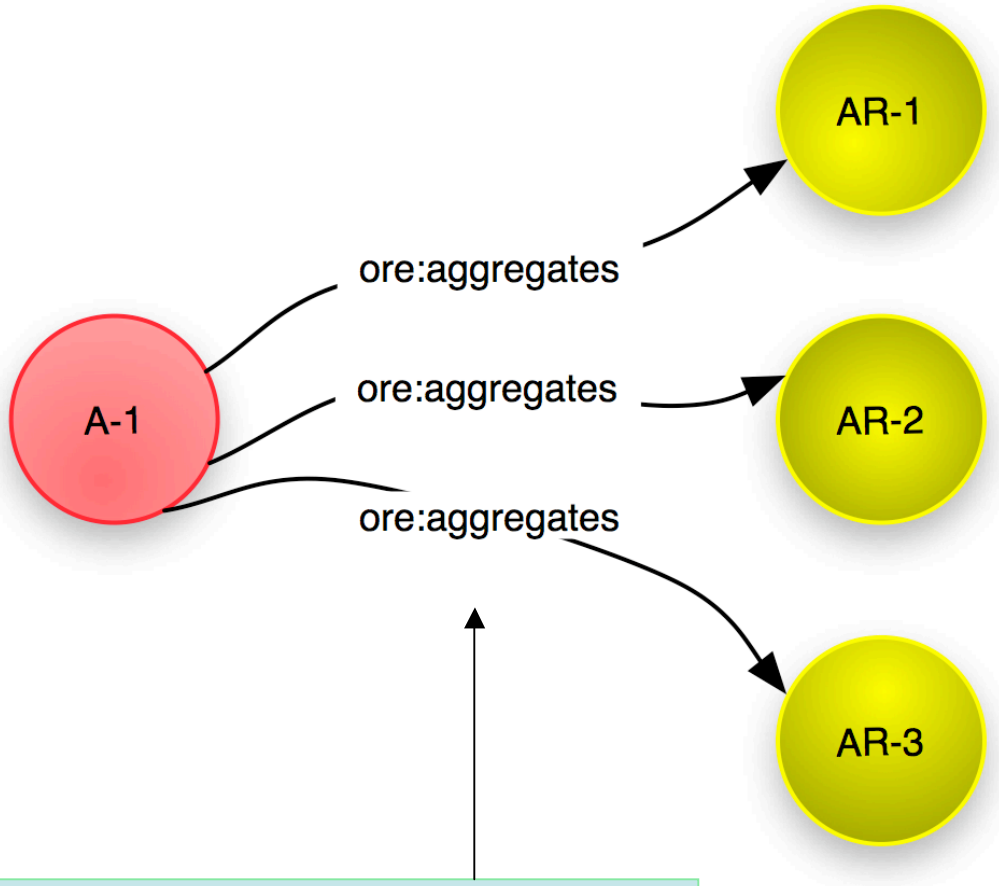
It starts with some resources that belong together



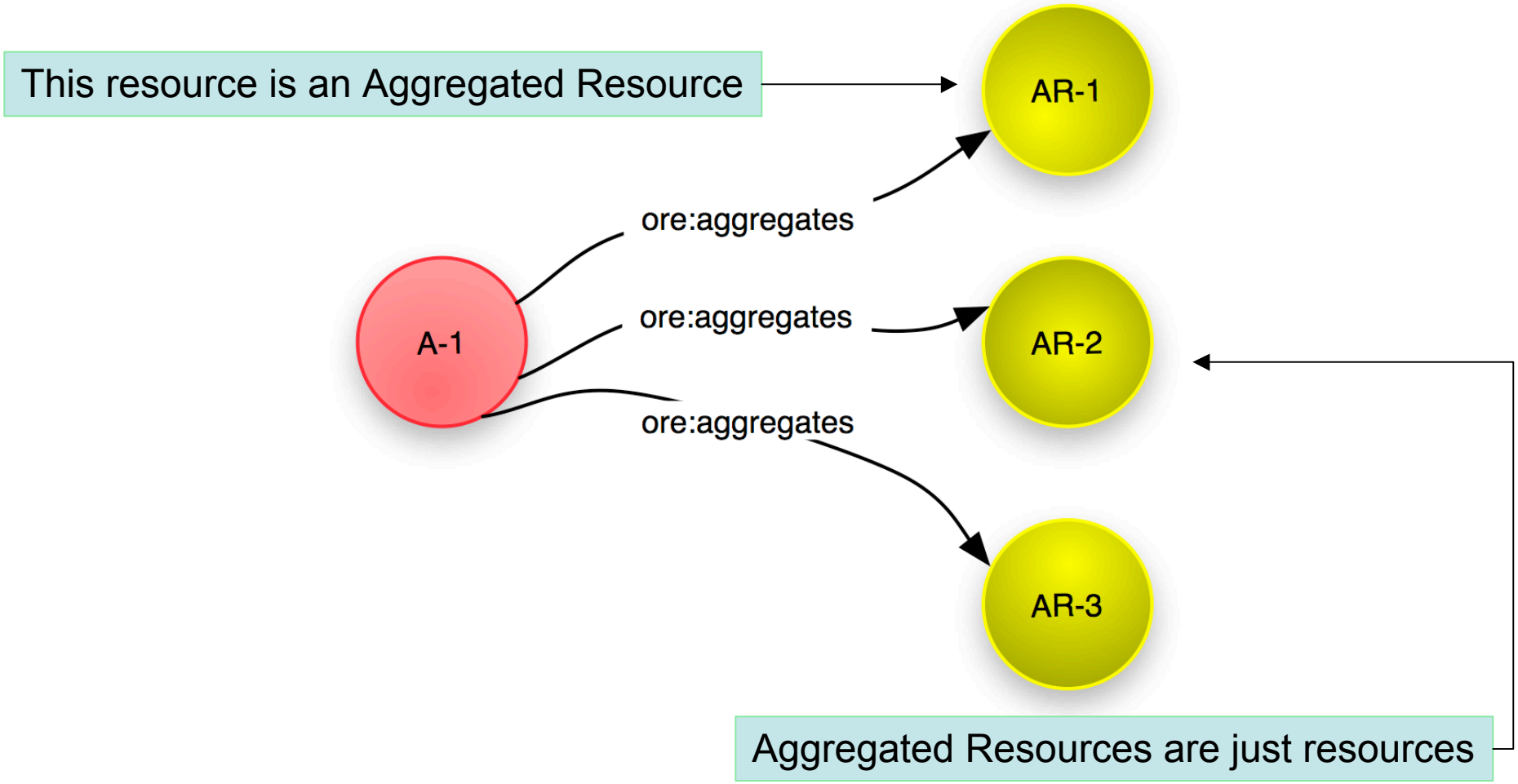
Introduce the Aggregation



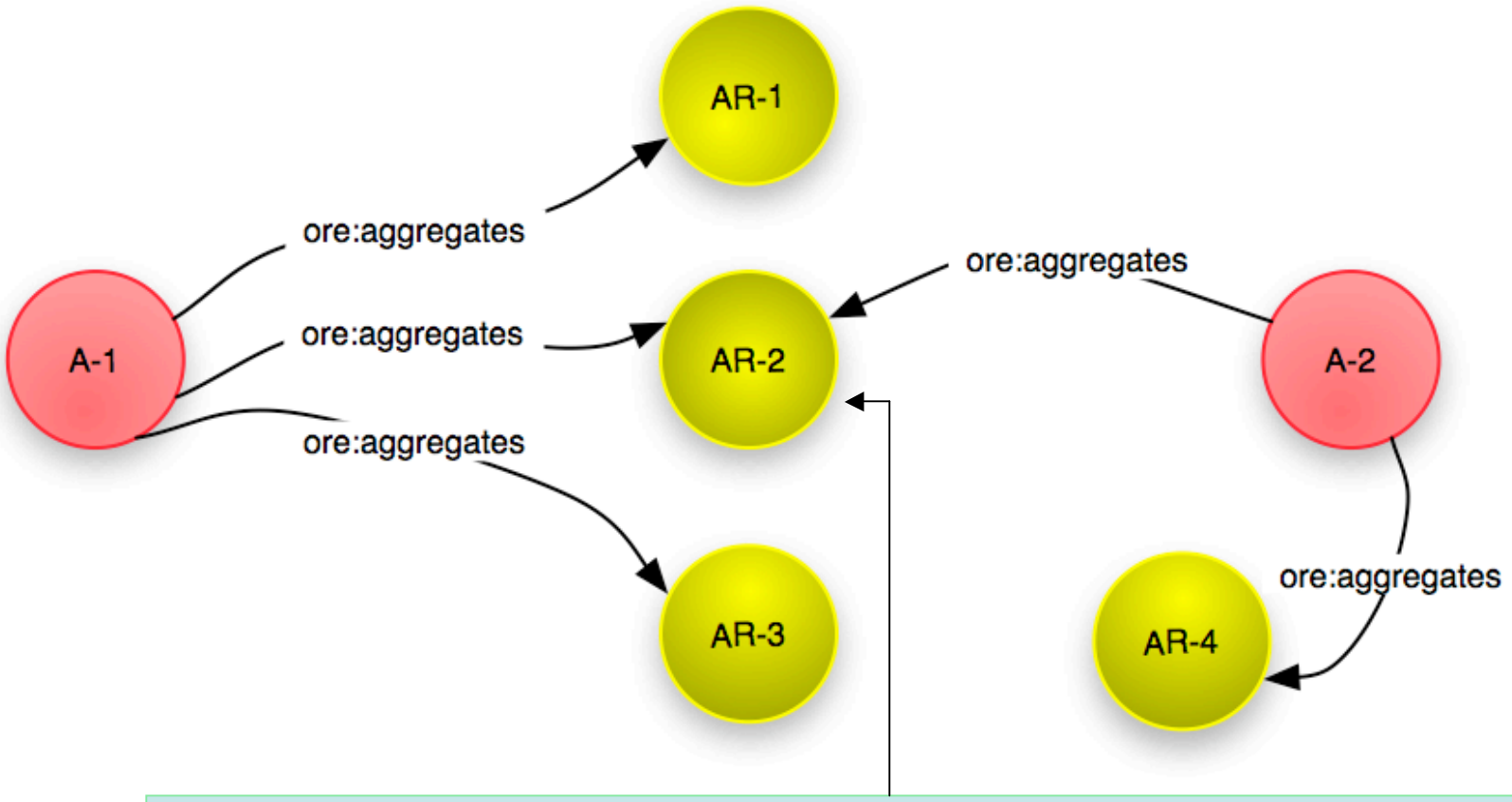
Express the `ore:aggregates` relationship



The `ore:aggregates` relationship

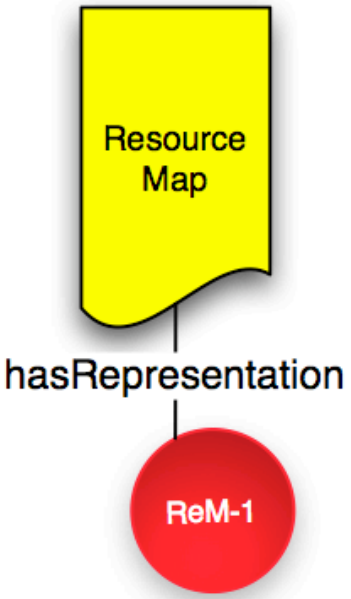


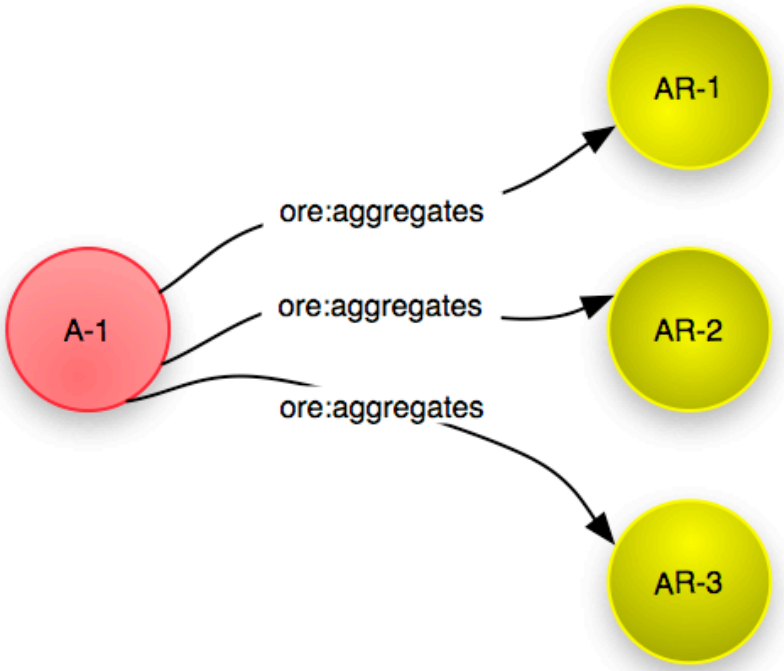
The `ore:aggregates` relationship



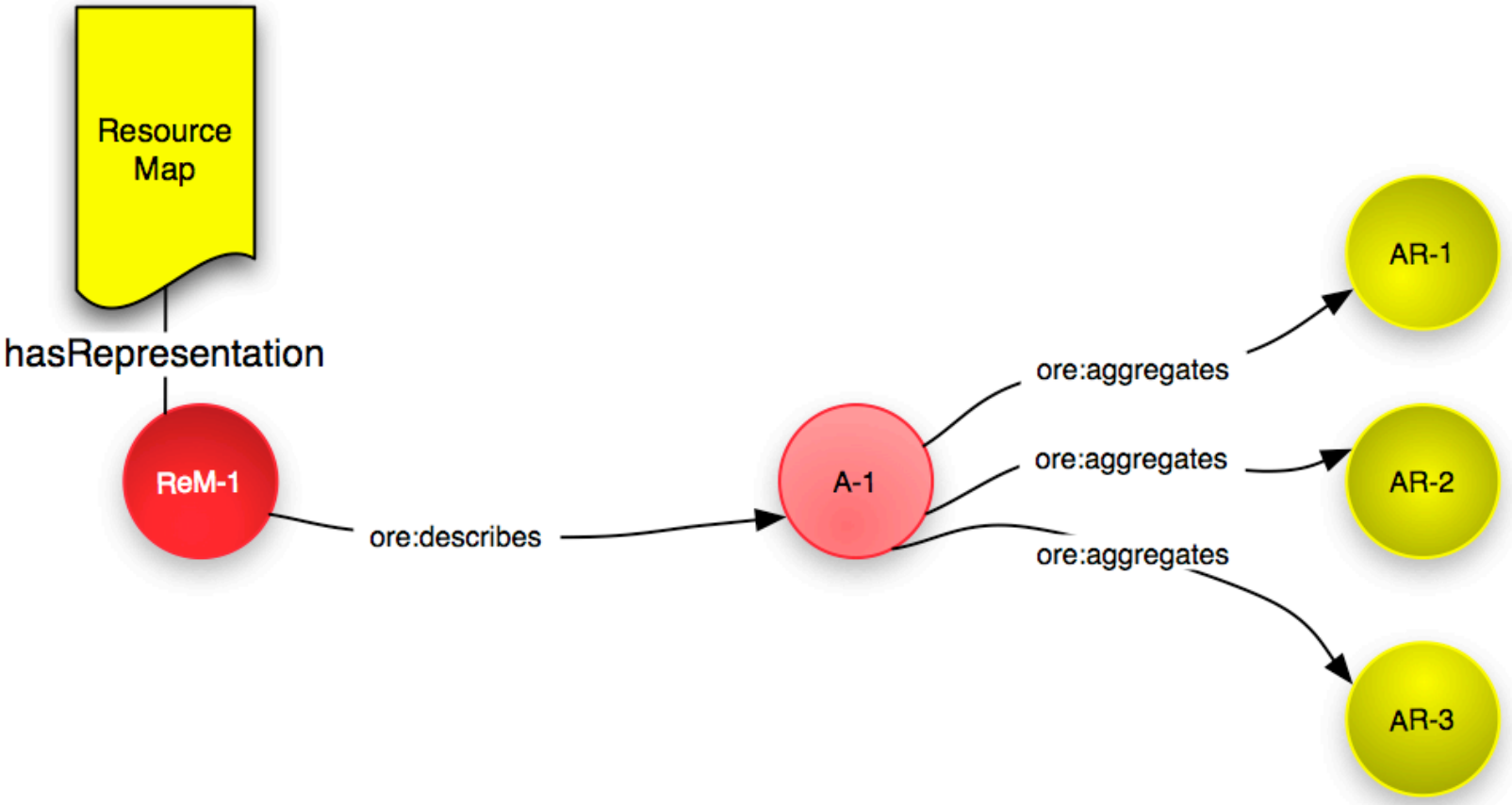
Aggregated Resources can be aggregated by multiple Aggregations

Introduce the Resource Map

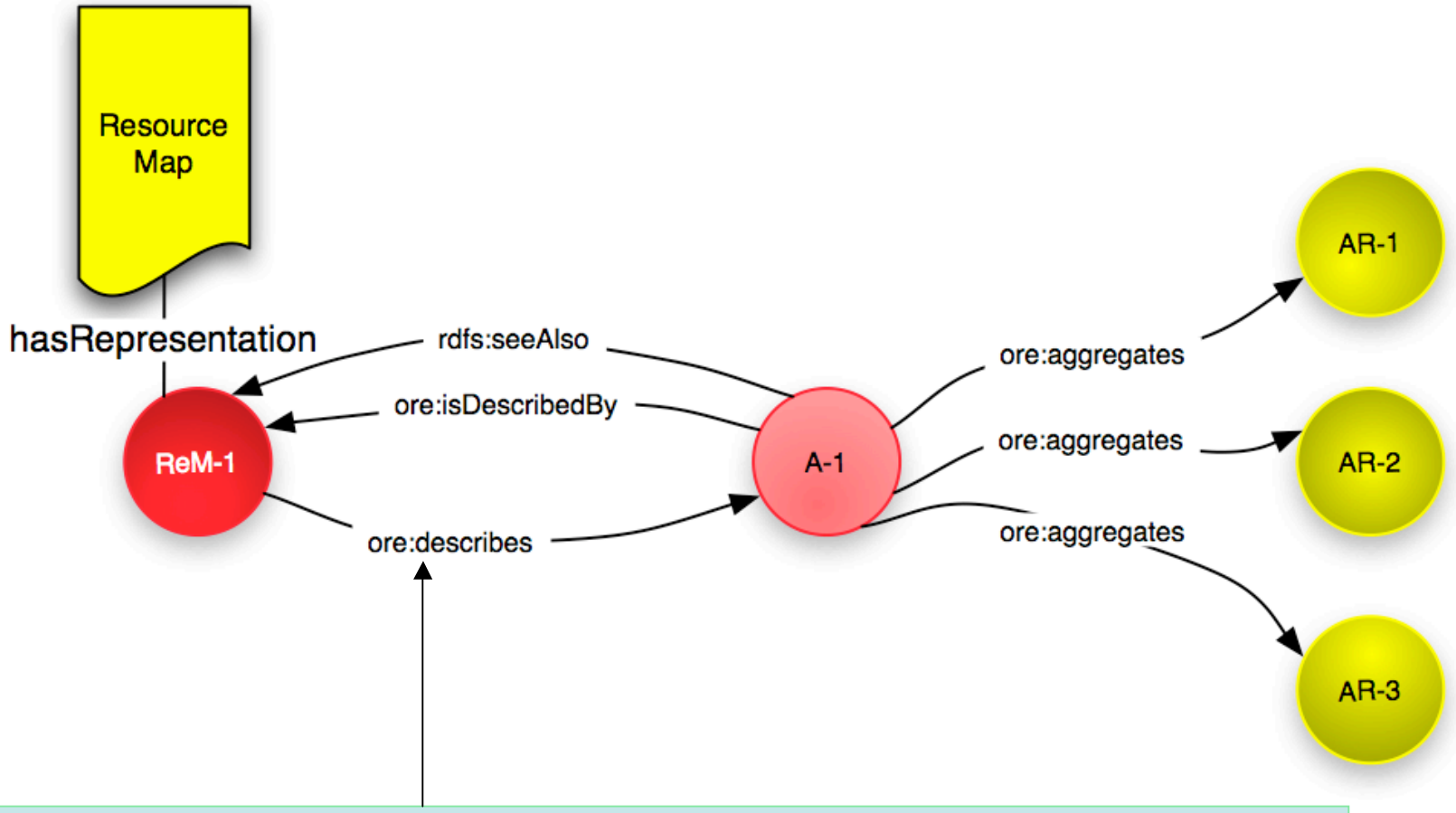




Express the `ore:describes` relationship

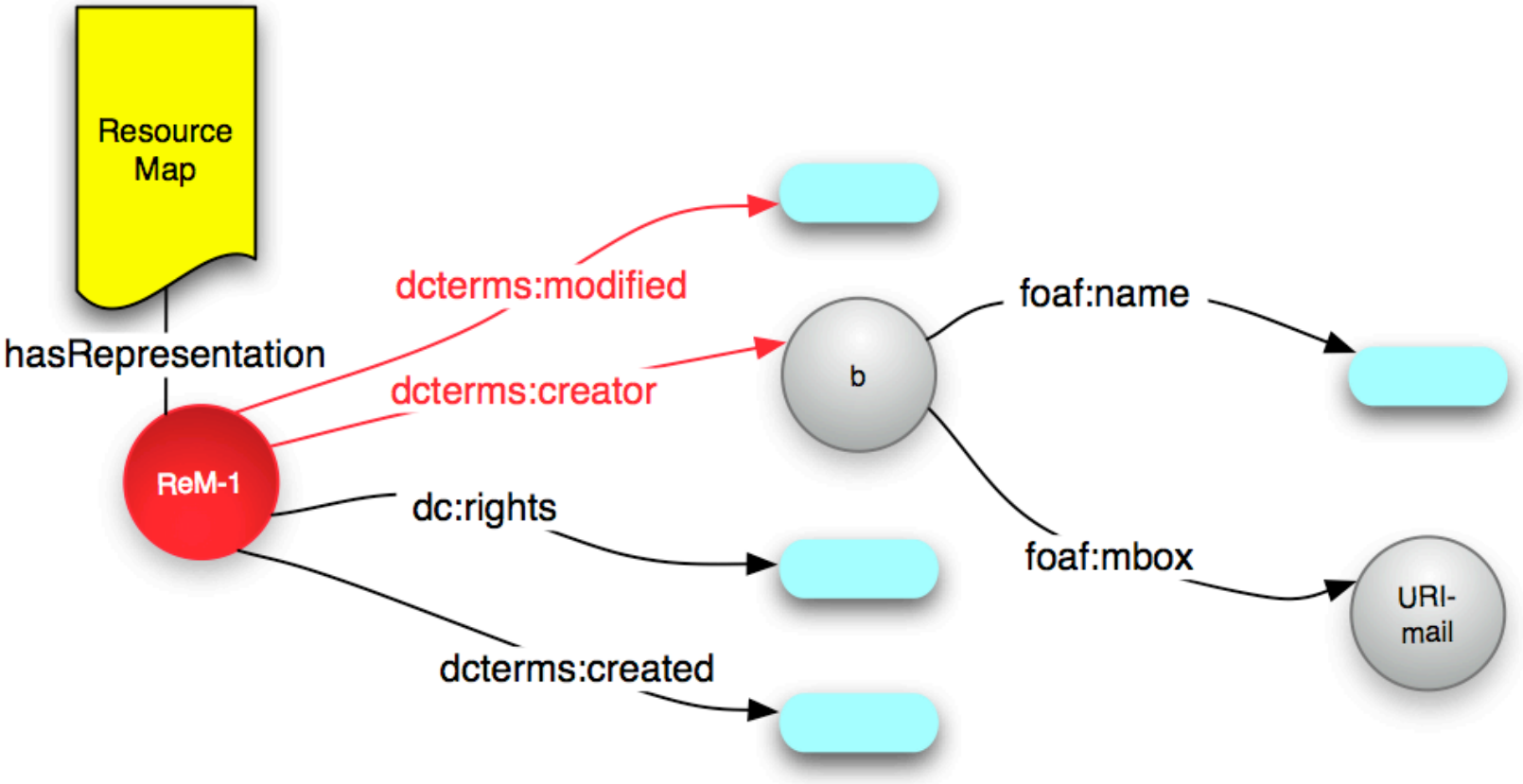


The `ore:isDescribedBy` relationship



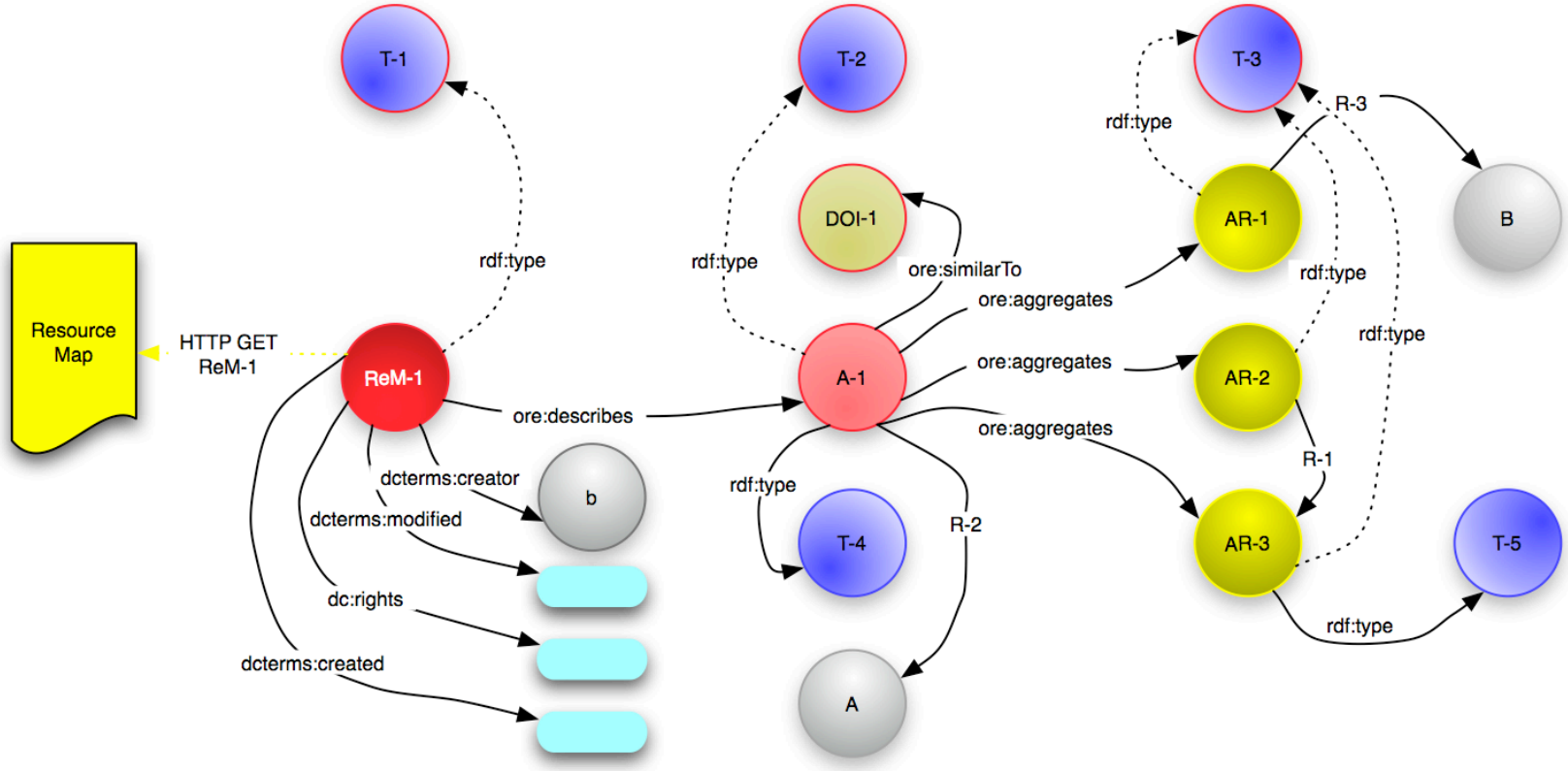
The inverse is `ore:isDescribedBy`; subproperty of `rdfs:seeAlso`

Express metadata about the Resource Map



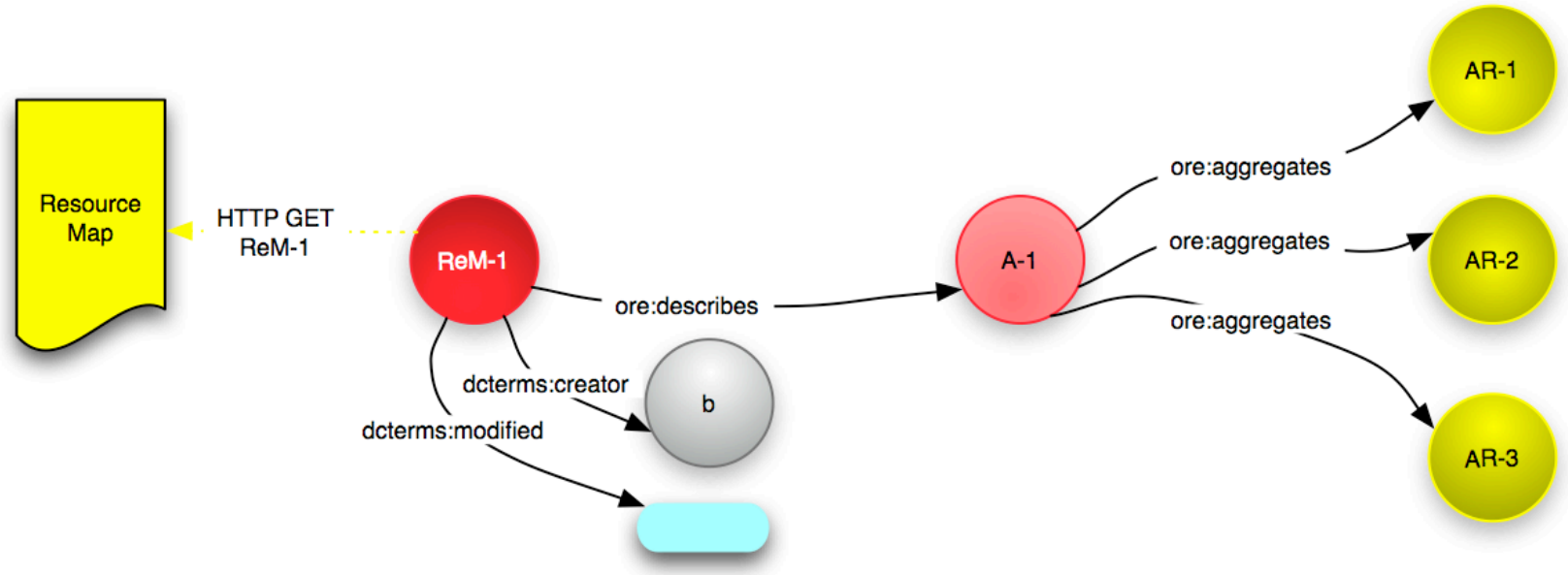
This corresponds to **metadata** from the Linked Data recommendations

A Resource Map can describe a lot …



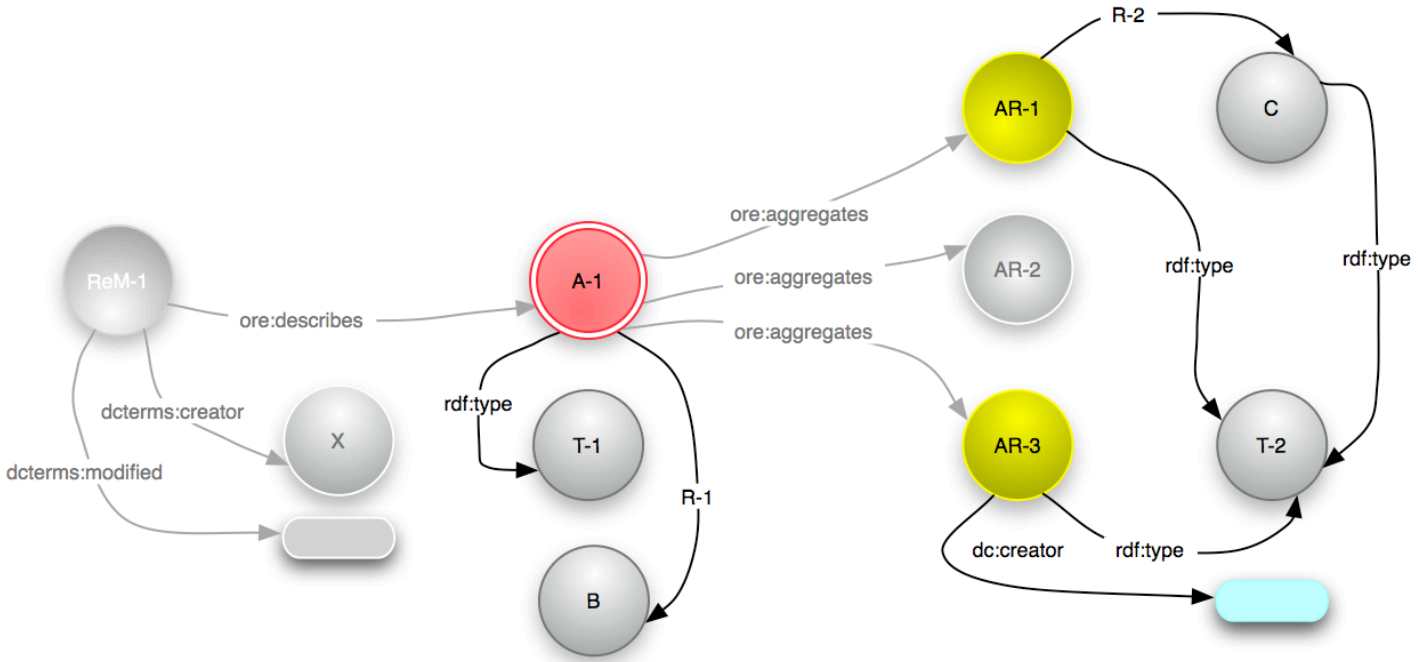
This corresponds to the **description**, **related descriptions**, **backlinks**, **metadata** from the Linked Data recommendations

But minimally it describes this …



This corresponds to the **description** (minimal), and **metadata** from the Linked Data recommendations

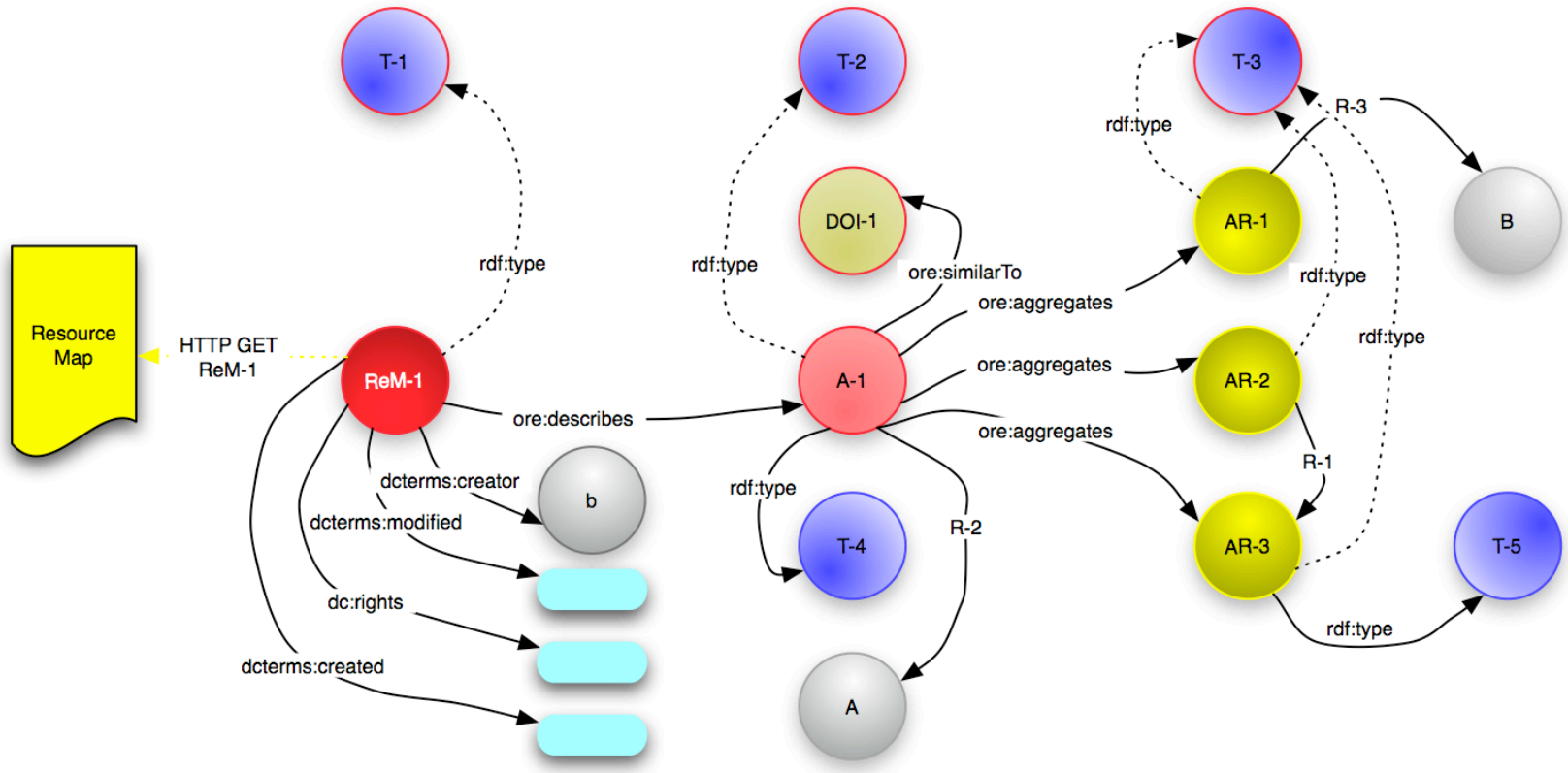
A lot about the Aggregation and the Aggregated Resources



<A-1>	<rdf:type>	<T-1>
<A-1>	<R-1>	<B>
<AR-1>	<rdf:type>	<T-2>
<AR-1>	<R-2>	<C>
<AR-3>	<rdf:type>	<T-2>
<C>	<rdf:type>	<T-2>

B	http://www.dlib.org
C	http://dx.doi.org/10.1023/A.1019213109274
A-1	http://www.dlib.org/dlib/february06/smith/aggregation
AR-1	http://www.dlib.org/dlib/february06/smith/02smith.html
AR-3	http://www.dlib.org/dlib/february06/smith/pg1-13.pdf
R-1	http://purl.org/dc/dcterms/isPartOf
R-2	http://purl.org/dc/dcterms/references
T-1	http://purl.org/eprint/type/JournalArticle
T-2	http://purl.org/dc/dcmitype/Text

A Resource Map can describe a lot …



but the graph expressed by the Resource Map must be **connected**

OAI Object Reuse and Exchange: Advanced 1

Multiple Resource Map Serializations

Authoritative Resource Maps

`e.g. HTTP 303`

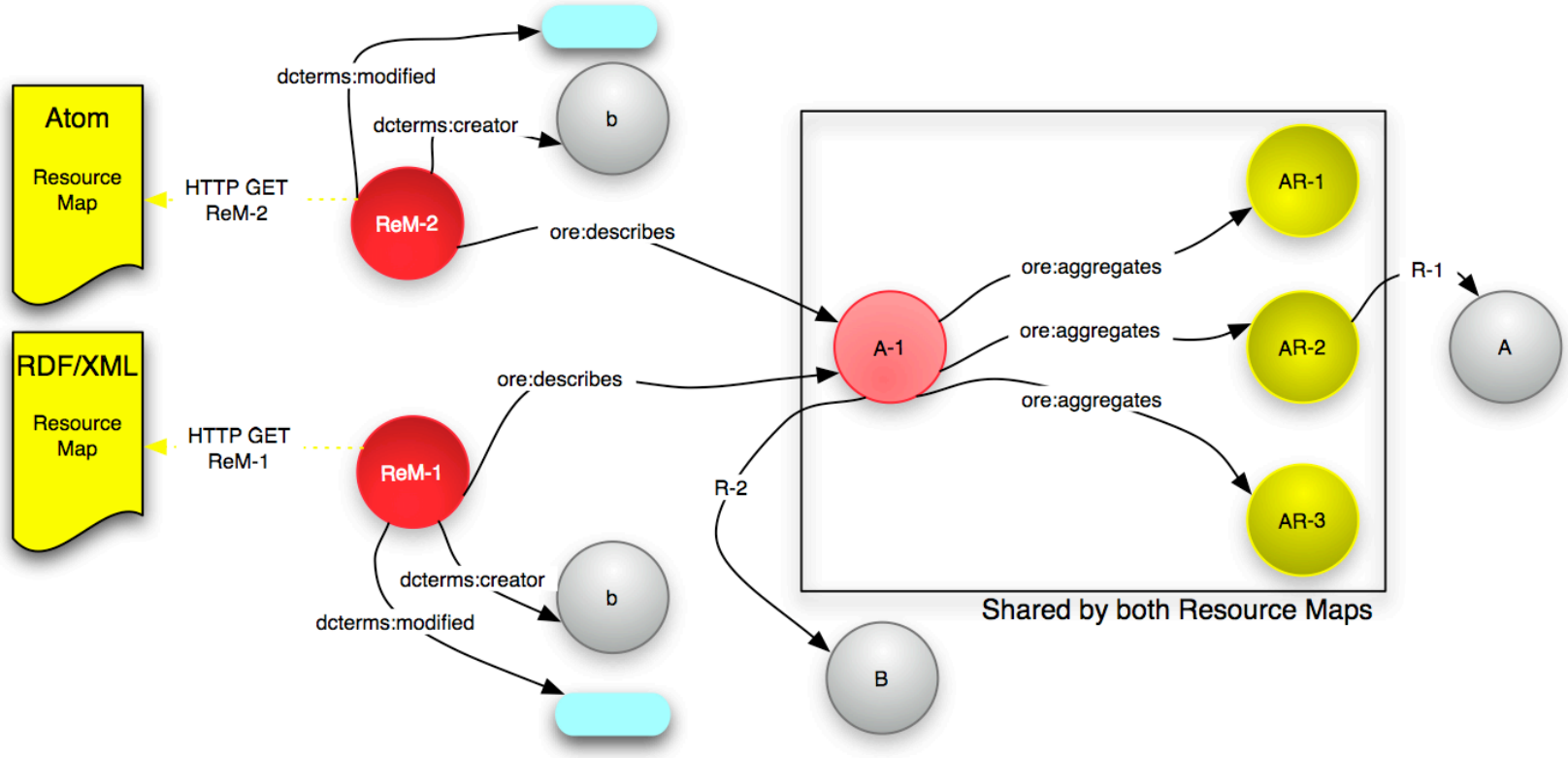
Discovery of Resource Maps

`ore:isDescribedBy`

Relationship between Aggregation and Resource Map

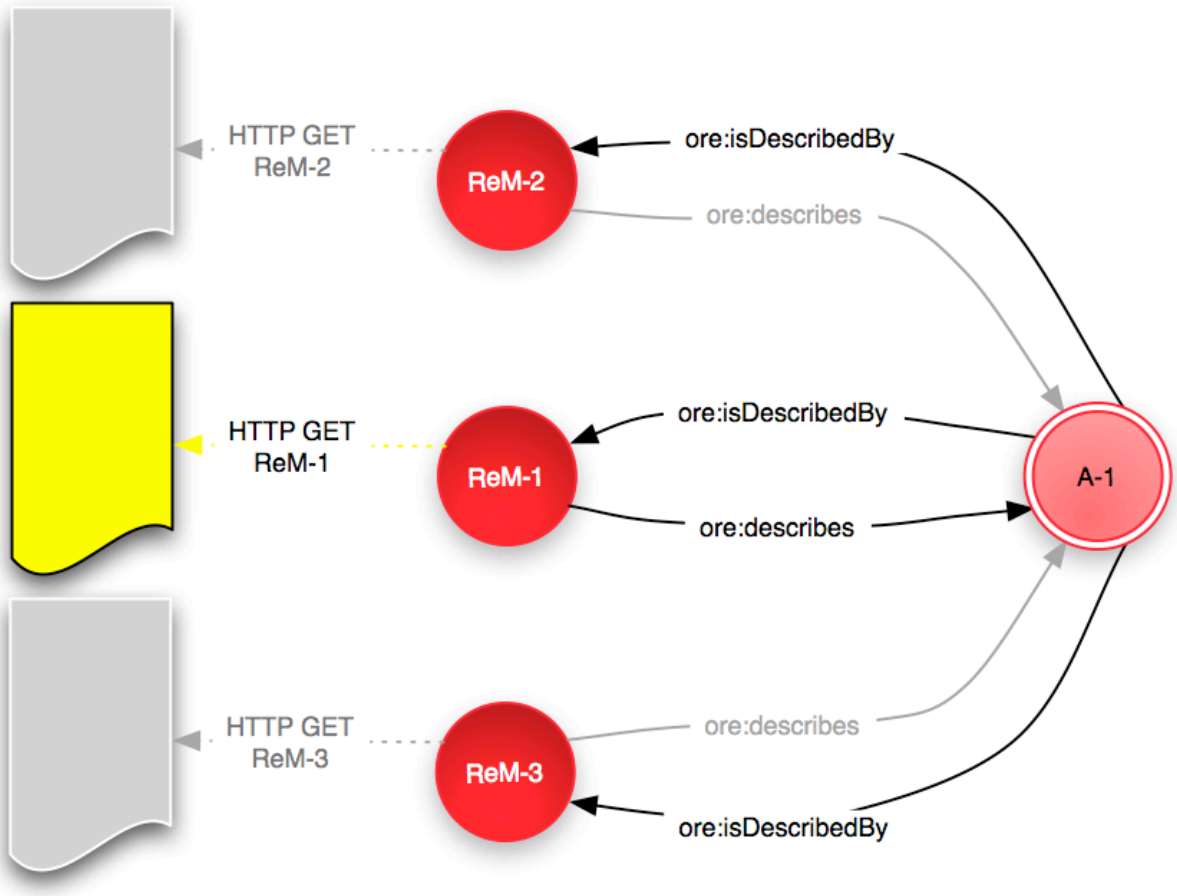
- An Aggregation is a Resource with a URI
- A Resource Map is a Resource with a URI
- A Resource Map asserts (identifies) and describes one Aggregation
  - A Resource is an Aggregation due to an assertion by (at least) one Resource Map
  - A Resource Map must have one representation


Multiple Resource Maps for an Aggregation; serializations



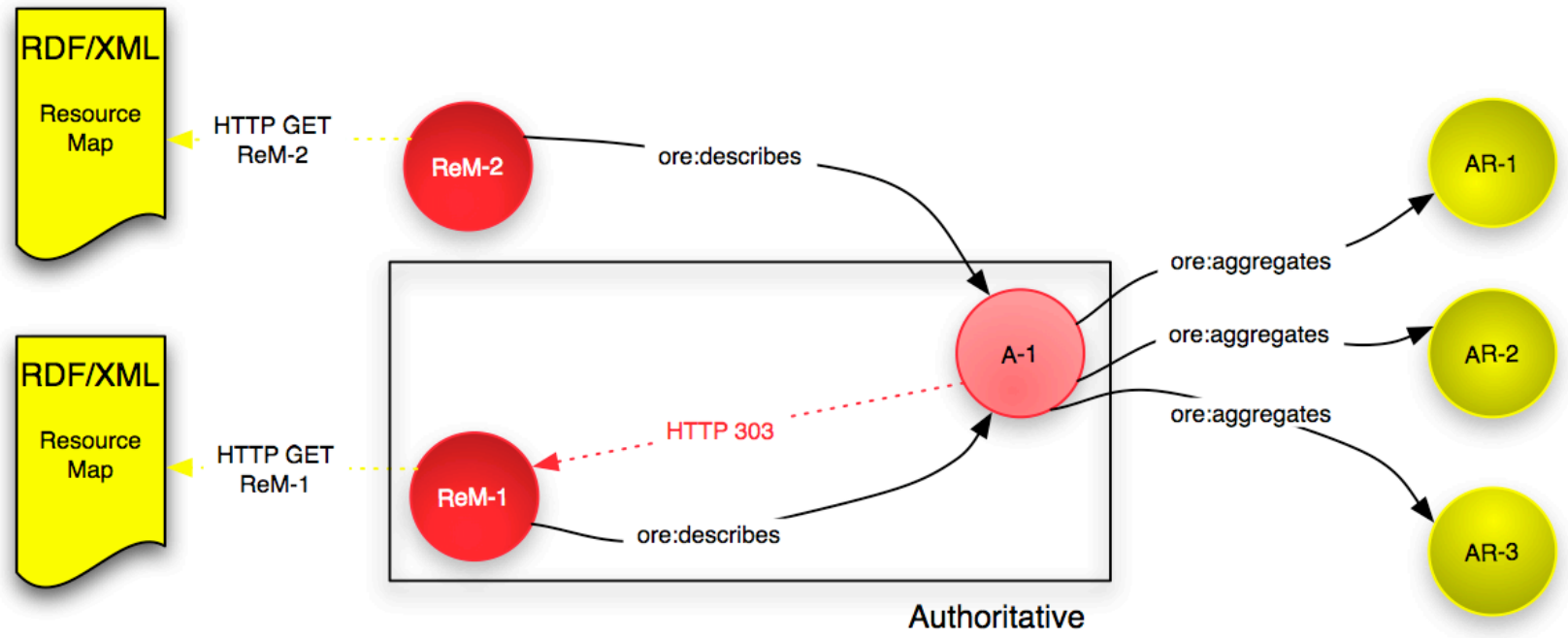
Aggregation Graph shared by both Resource Maps. Also Proxies shared (later).

Multiple Resource Maps for an Aggregation; discovery



Use `ore:isDescribedBy` to facilitate discovery of other Resource Maps

Multiple Resource Maps for an Aggregation; authoritative



Authoritative: dereference of URI of Aggregation leads to Resource Map

Authoritative and. Non-Authoritative Resource Maps

- Authoritative
  - Created by same authority (usually)
  - Must be minimally equivalent (same Aggregated Resources and Proxies)
  - Should assert mutual existence (`ore:isDescribedBy`)
- Non-authoritative
  - Best practice is to not create them
  - Assert your own Aggregation instead
  - Use `rdfs:seeAlso` to assert relationship between two Aggregations

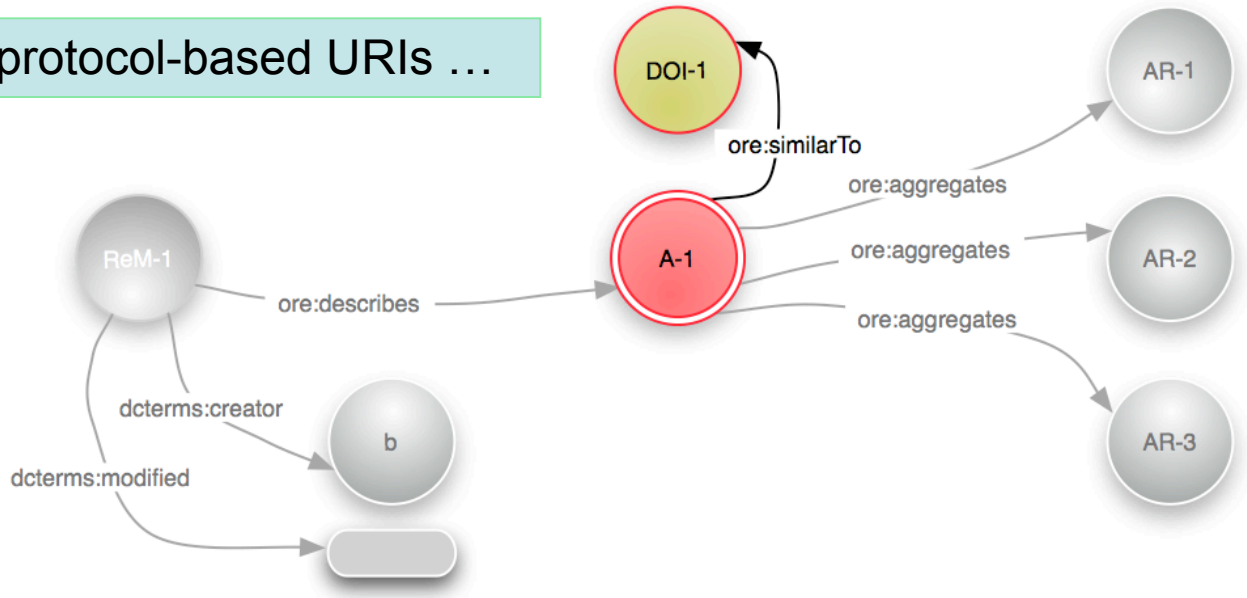
OAI Object Reuse and Exchange: Advanced 2

Expressing non-protocol-based URIs

`ore:similarTo`

The `ore:similarTo` relationship

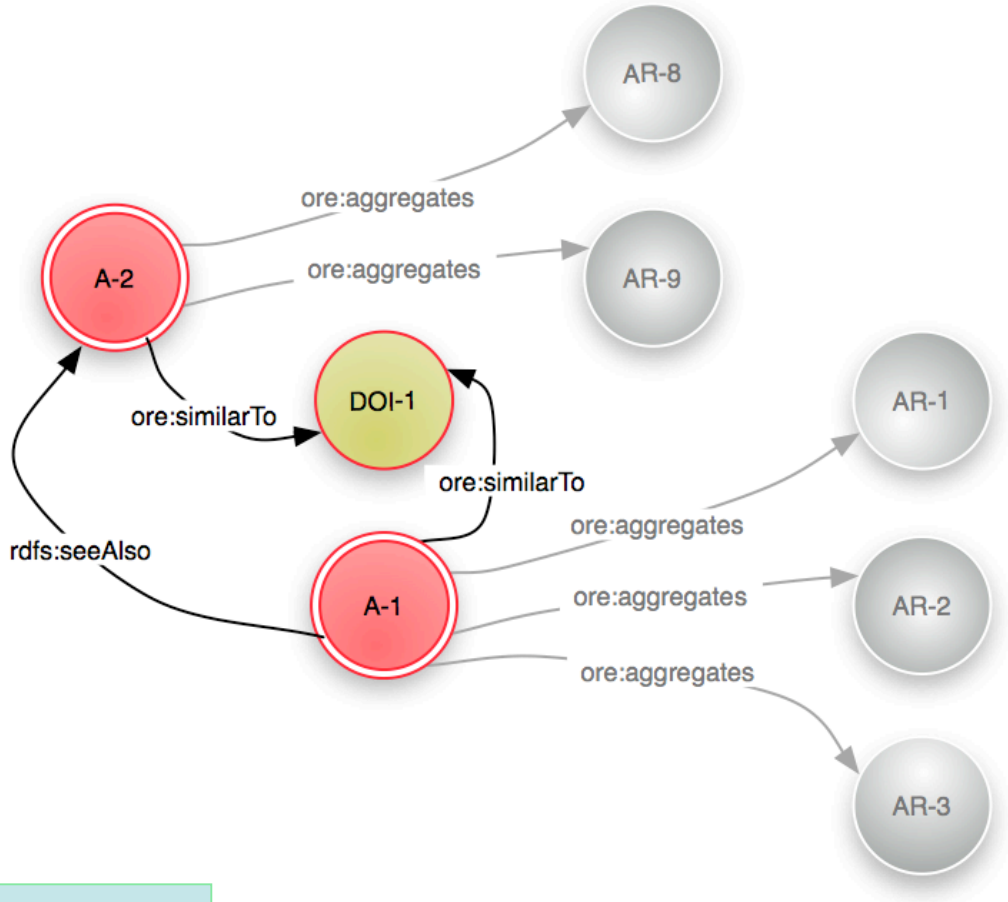
To express non-protocol-based URIs …



<A-1>	<ore:similarTo>	<DOI-1>

A-1	http://www.dlib.org/dlib/february06/smith/aggregation
DOI-1	info:doi/10.1045/february-2006-smith
ore:similarTo	http://www.openarchives.org/ore/terms/similarTo

The `ore:similarTo` relationship



DOI-1 connects the graphs

The bigger message re URI-A

- Mint a new URI-A for an Aggregation
  - Use a HTTP URI
- And mint new URI-Rs for Resource Maps that describe the Aggregation
  - Use HTTP URIs
- Do not overload:
  - The DOI
  - The splash page URI-S

  by turning them into URI-A.
- Rather express relationships between those URIs and URI-A:
  - `URI-A ore:aggregates URI-S`
  - `URI-S rdf:type info:eu-repo/semantics/humanStartPage`
  - `URI-A ore:similarTo DOI-1`


OAI Object Reuse and Exchange: Advanced 3

Aggregated Resource member of another Aggregation

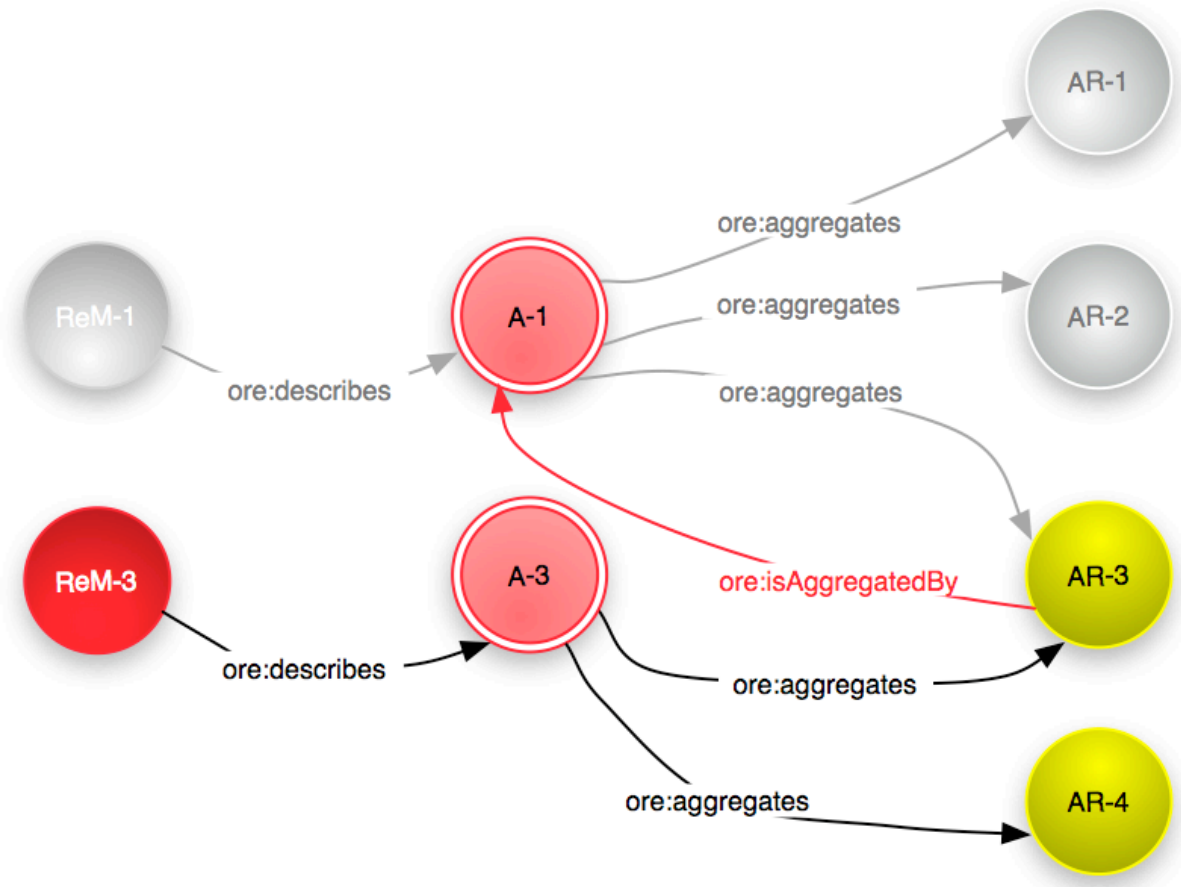
`ore:isAggregatedBy`

Aggregated Resource is an Aggregation

`ore:isDescribedBy`

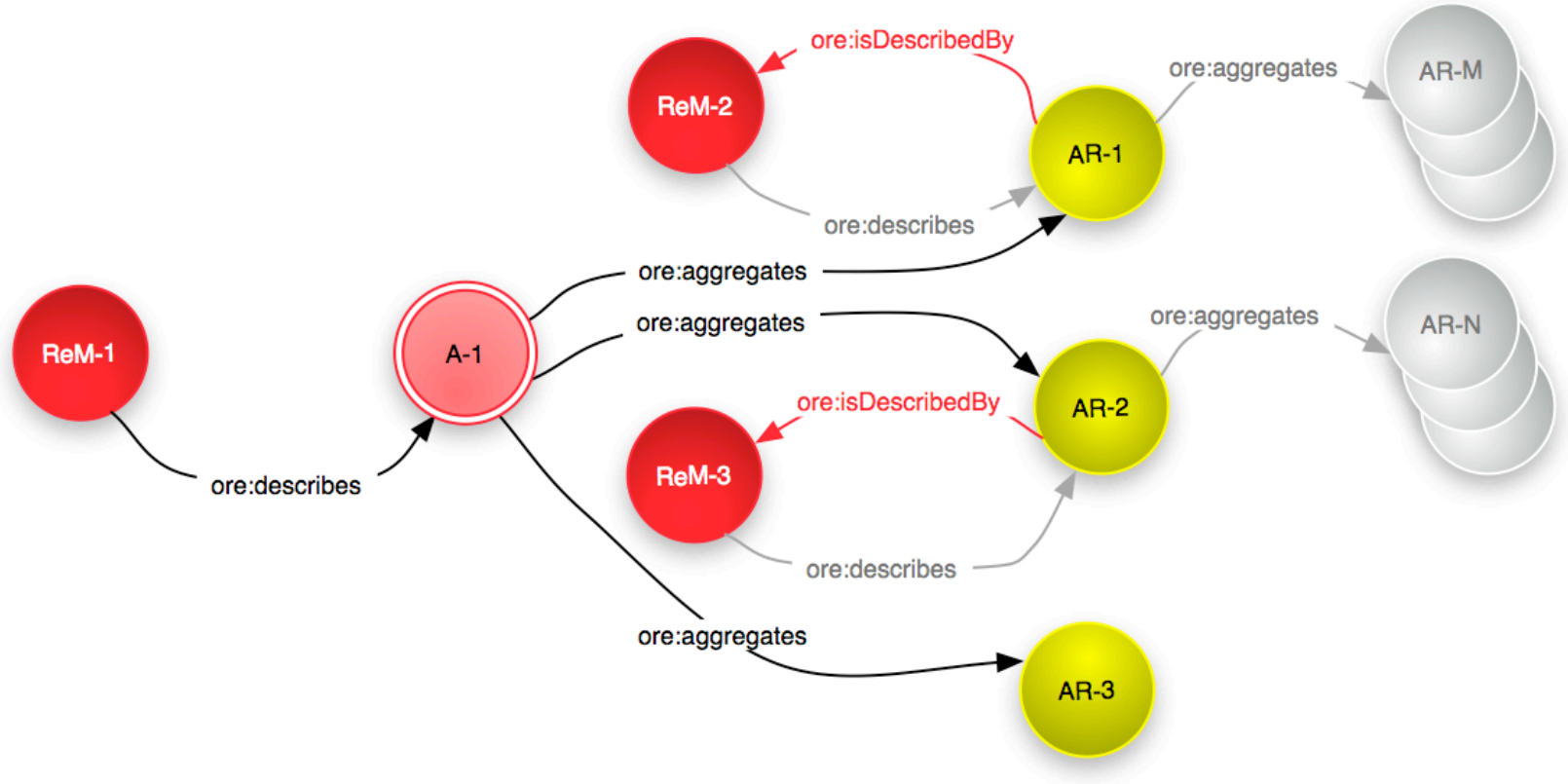
Los Alamos
NATIONAL LABORATORY

A resource is an Aggregated Resource is another Aggregation



Use `ore:isAggregatedBy` to express membership of another Aggregation

An Aggregated Resource is itself an Aggregation



Use `ore:isDescribedBy` to point at a Resource Map that describes that Aggregation

OAI Object Reuse and Exchange: Advanced 4

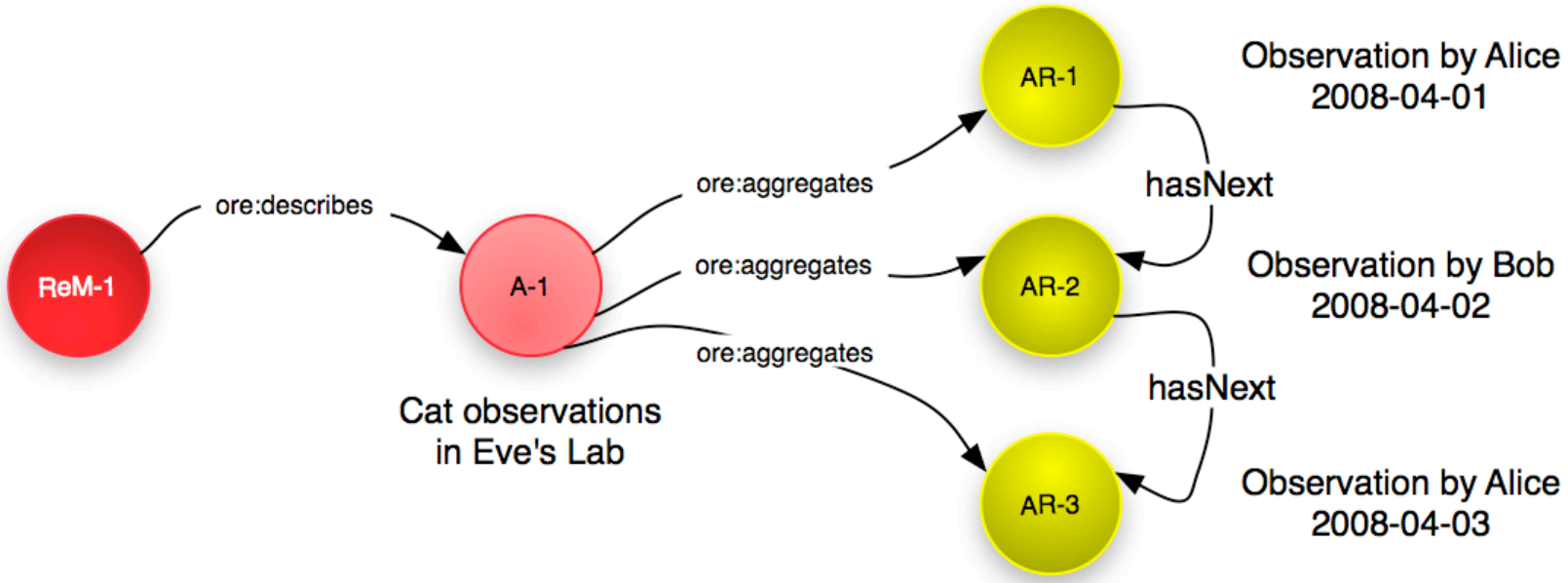
Proxy: Aggregated Resource in Context of an Aggregation

`ore:isProxyFor`

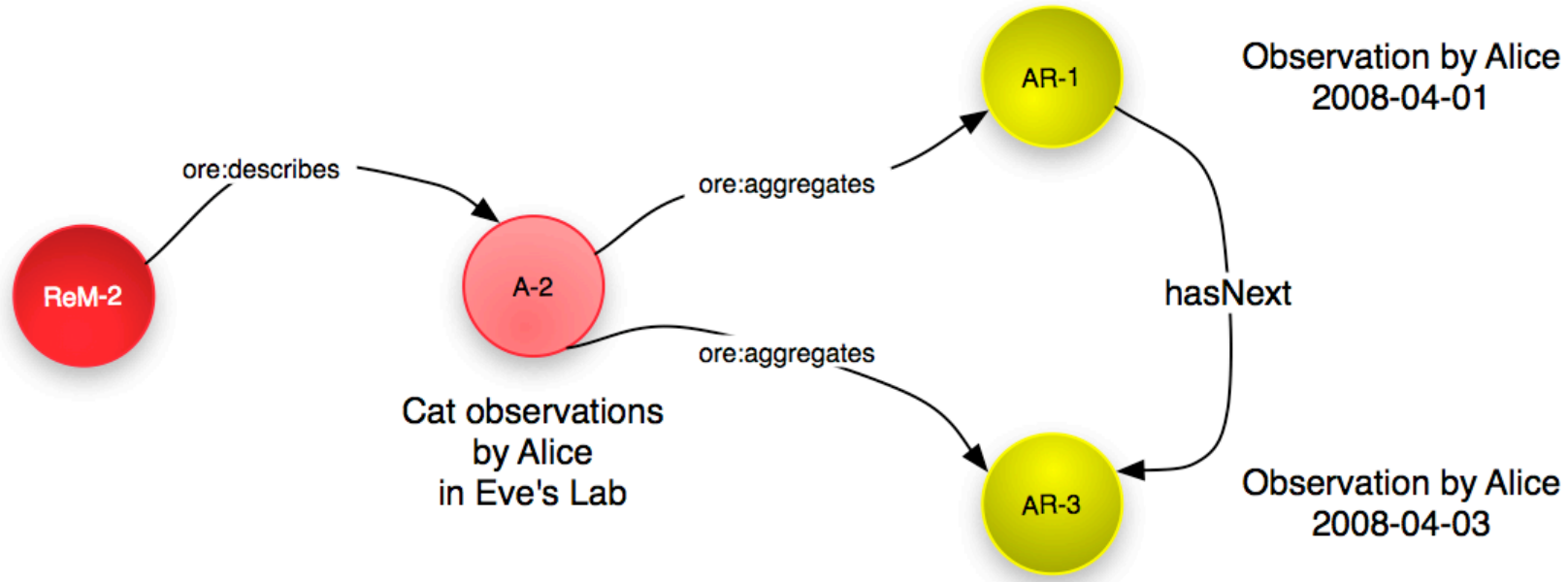
`ore:isProxyIn`

Alice and Bob observe cats in Eve's Lab

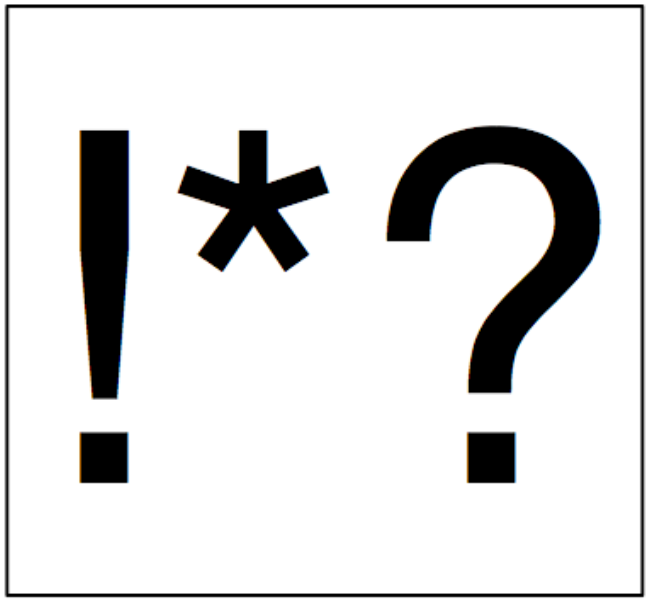
Alice and Bob observe cats in Eve's Lab



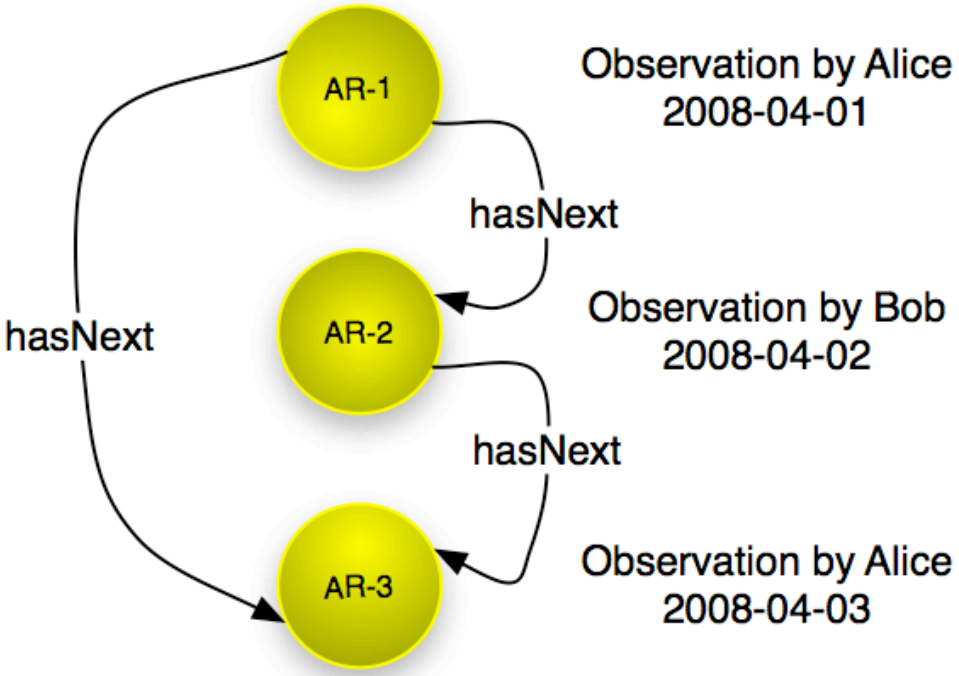
Alice observes cats in Eve's Lab



An agent merges information and gets confused



What the heck!



What did we mean by hasNext?

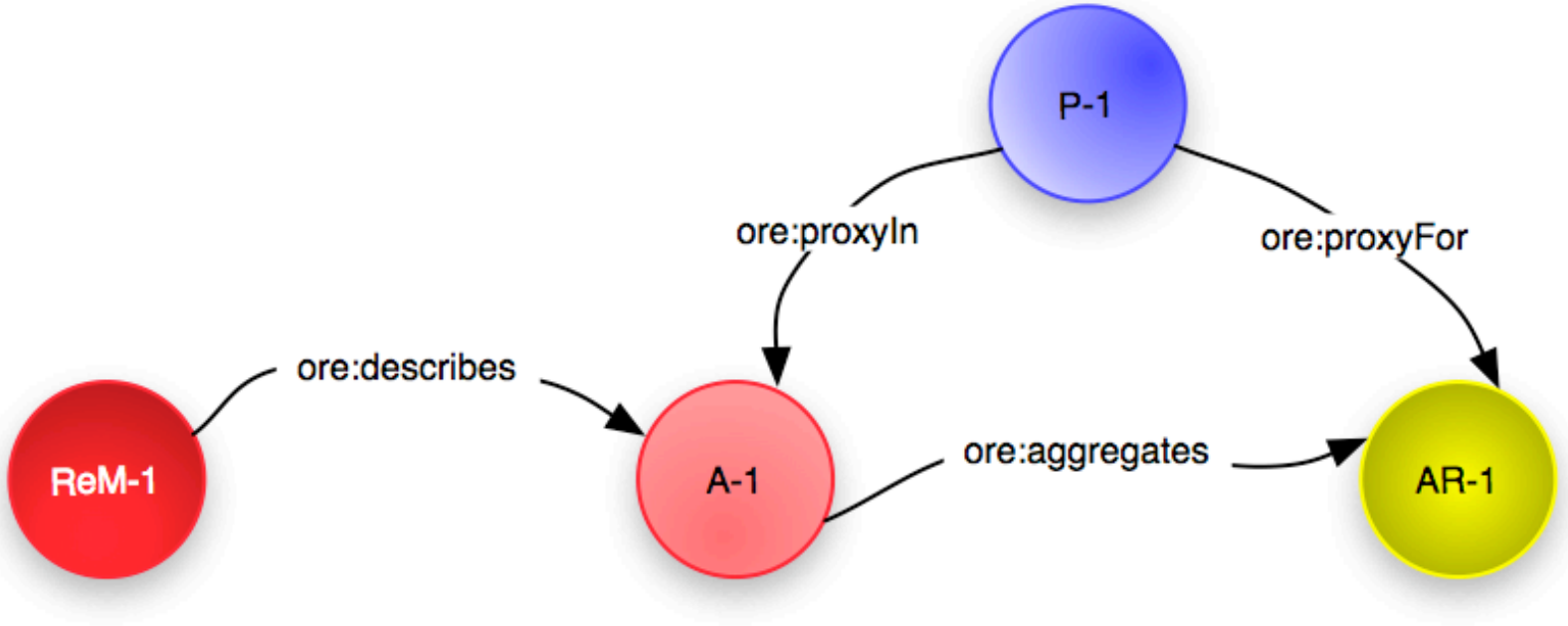
- Resource Map 1: Bob's observation on 2008-04-02 is the next observation after Alice's observation on 2008-04-01 in the sequence of observations in Eve's Lab

- Resource Map 2: Alice's observation on 2008-04-03 is the next observation after her observation on 2008-04-01 in the sequence of Alice's observations in Eve's Lab

Modeling a Resource in the Context of an Aggregation: Proxy

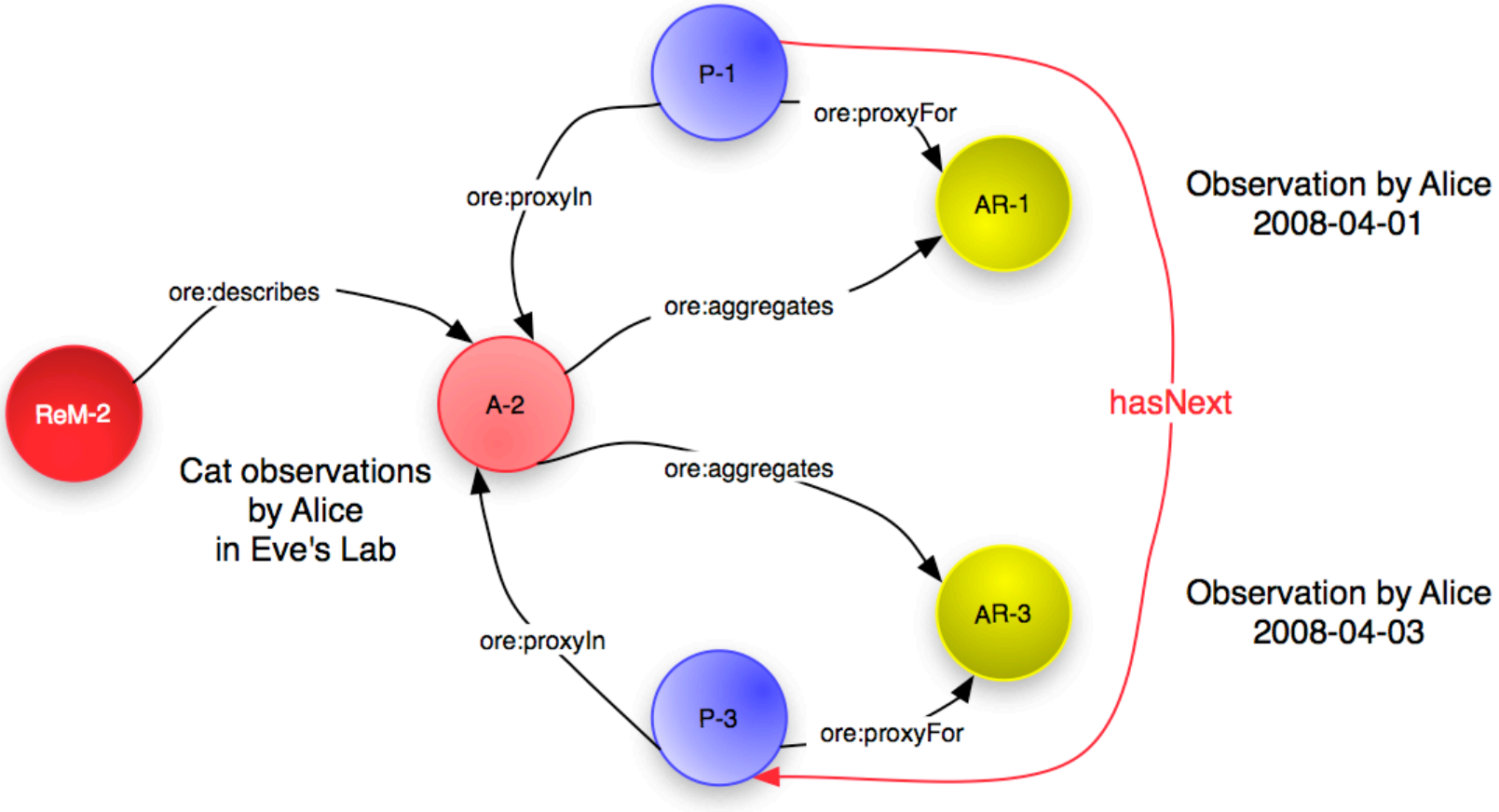
- Two components:
  - The (Aggregated) Resource
  - The context in which it is aggregated, i.e. the Aggregation
- In the Web Architecture, a new concept needs a new resource (and hence URI): we named it the Proxy

Modeling a Resource in the Context of an Aggregation: Proxy



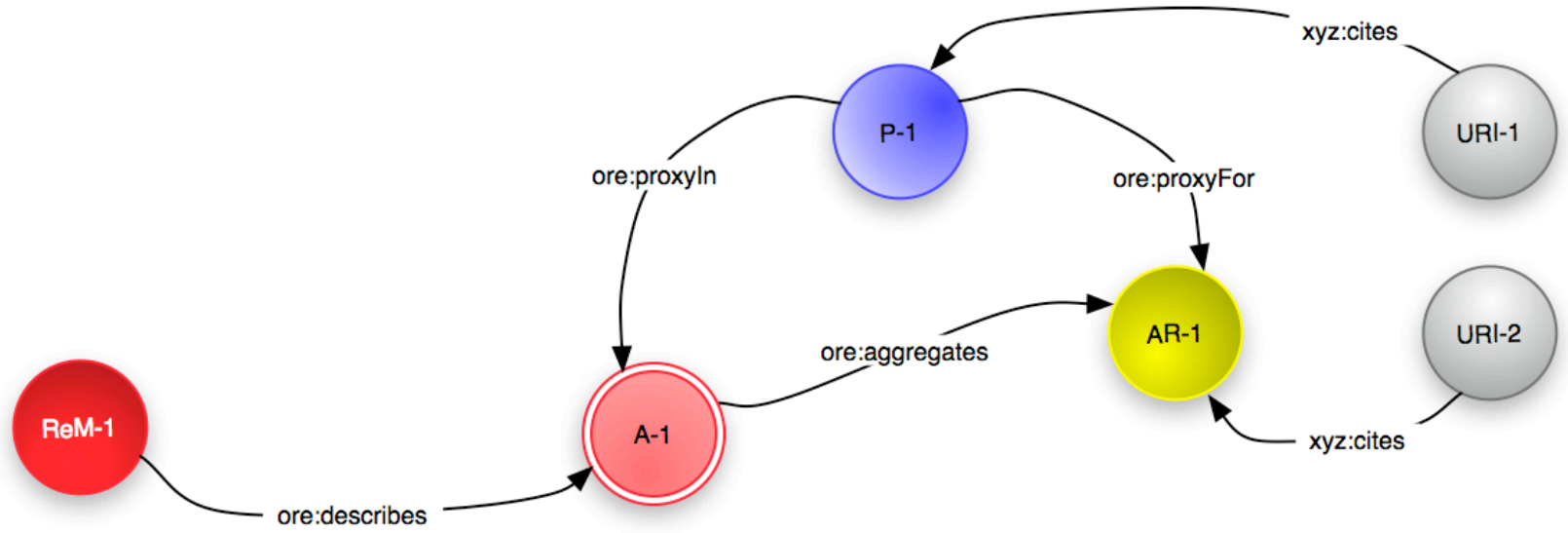
`ore:proxyFor` and `ore:proxyIn` to introduce a Proxy for an Aggregated Resource

Alice's observations in context



`hasNext` expressed as a relationship between Proxies

Citation to a resource in a specific context



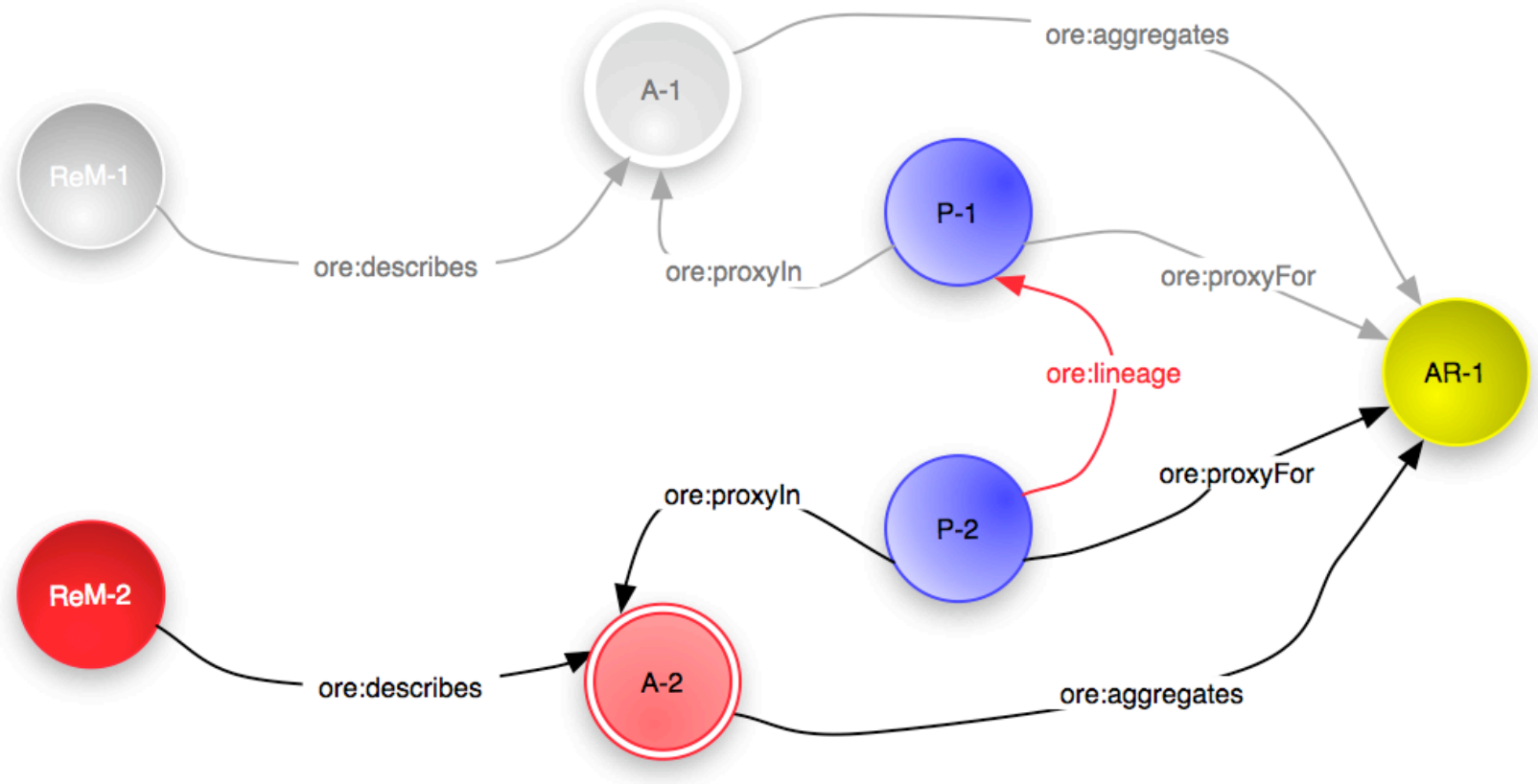
OAI Object Reuse and Exchange: Advanced 5

Lineage of an Aggregated Resource

`ore:lineage`

Los Alamos
NATIONAL LABORATORY

An Aggregated Resource *originated* in another Aggregation

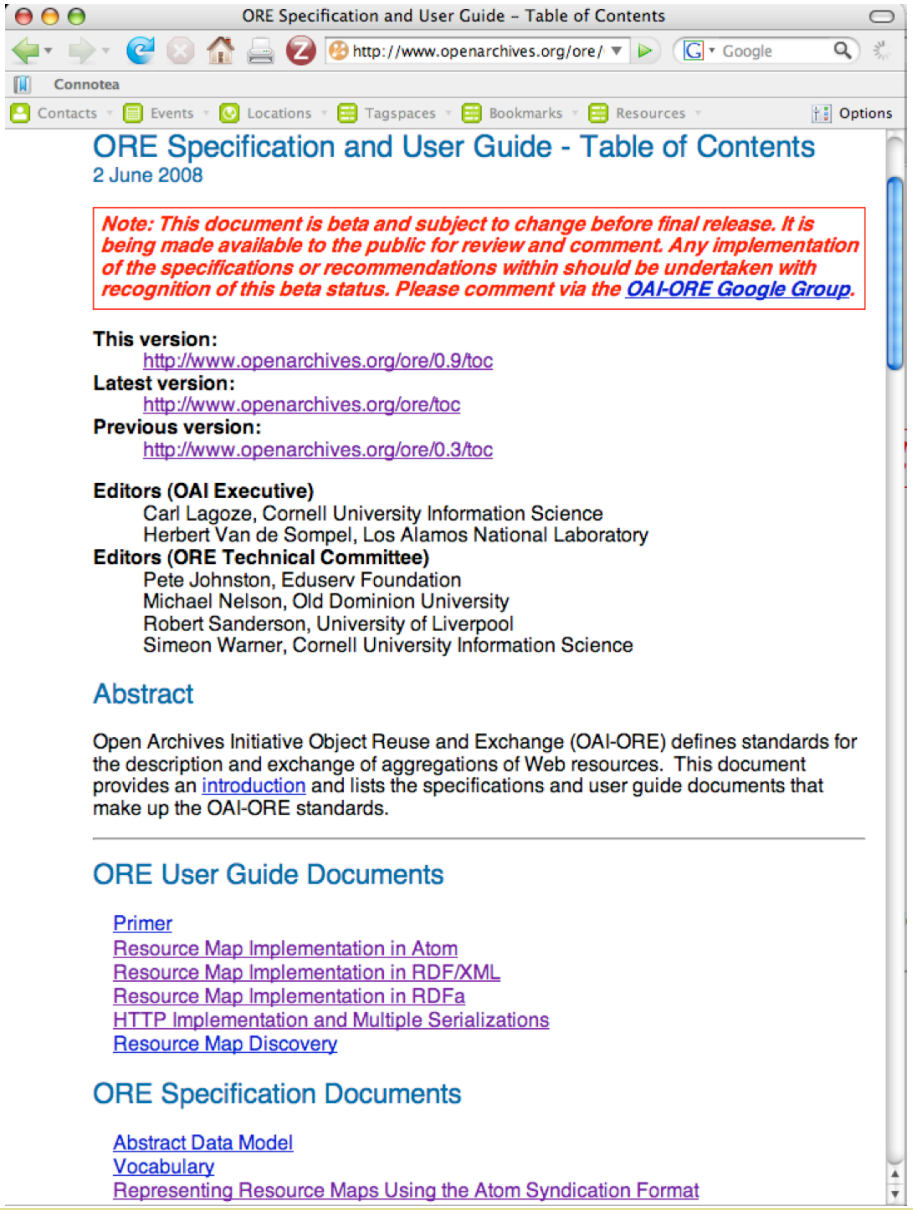


`ore:lineage` is a relationship between Proxies

OAI Object Reuse and Exchange

The ORE Documents
Expected Changes

- Beta version released June 2nd 2008
- Feedback very welcome via ORE Google Group
- Version 1.0 expected 09/2008



http://www.openarchives.org/ore/toc

The ORE documents: expected changes

- ORE Primer
  - Expect a major rewrite that will not overload readers with big concepts
- Atom Resource Maps
  - Expect a major change in the Atom serialization
  - Expect a major rewrite
- RDF/XML Resource Maps
- RDFa Resource Maps
- HTTP implementation
  - Expect clearer examples, and clear connection to Cool URIs for Semantic Web
- Discovery of Resource Maps
  - Expect changes related to the changes in Atom serialization
- Data Model
  - Expect removal/downplay of Named Graph stuff
- ORE Profile of Atom: gone


OAI Object Reuse and Exchange

Atom Serialization

The Atom serialization: expected changes

- See http://www.openarchives.org/ore/documents/atom_revision_20080801.html
- 2 core changes:
  - Describe an Aggregation at the level of an Atom Entry, not Feed.
  - Do not overload existing Atom semantics/relationships; rather introduce ORE-specific ones, e.g.

    ```
    <link rel="ore:aggregates" href="AR-1">
    ```
- Example

OAI Object Reuse and Exchange

Playing ORE in two worlds

Interoperability Stacks

Atom profiles, APP, Special-purpose APIs	Vocabularies, SPARQL
Feed technologies, RSS, Atom	RDF, RDF serializations
HTTP URI	HTTP URI
Web 2.0	**Semantic Web; Linked Data**

Interoperability Stacks

ORE add-ons for Atom	Atom profiles, APP, special-purpose APIs	Vocabularies, SPARQL	ORE terms, dcterms, foaf
Atom ReM	Feed technologies, RSS, Atom	RDF, RDF serializations	RDF-based data model; HTTP 303; RDF/XML ReM, RDFa ReM
HTTP URI for Aggregation, Resource Map, Proxies	HTTP URI	HTTP URI	HTTP URI for Aggregation, Resource Map, Proxies
ORE	**Web 2.0**	**Semantic Web; Linked Data**	**ORE**

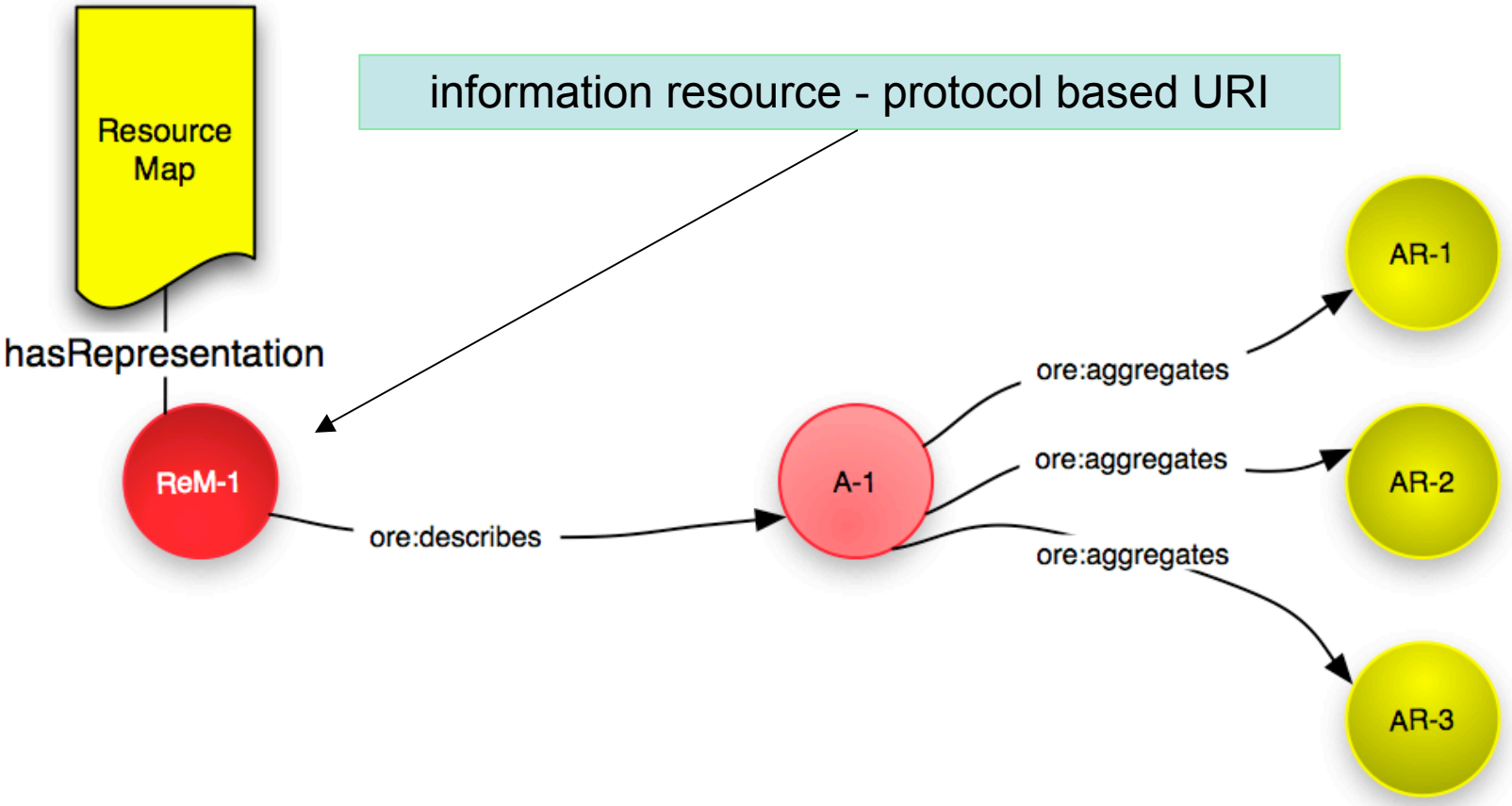
Los Alamos NATIONAL LABORATORY

OPEN ARCHIVES

OAI Object Reuse and Exchange

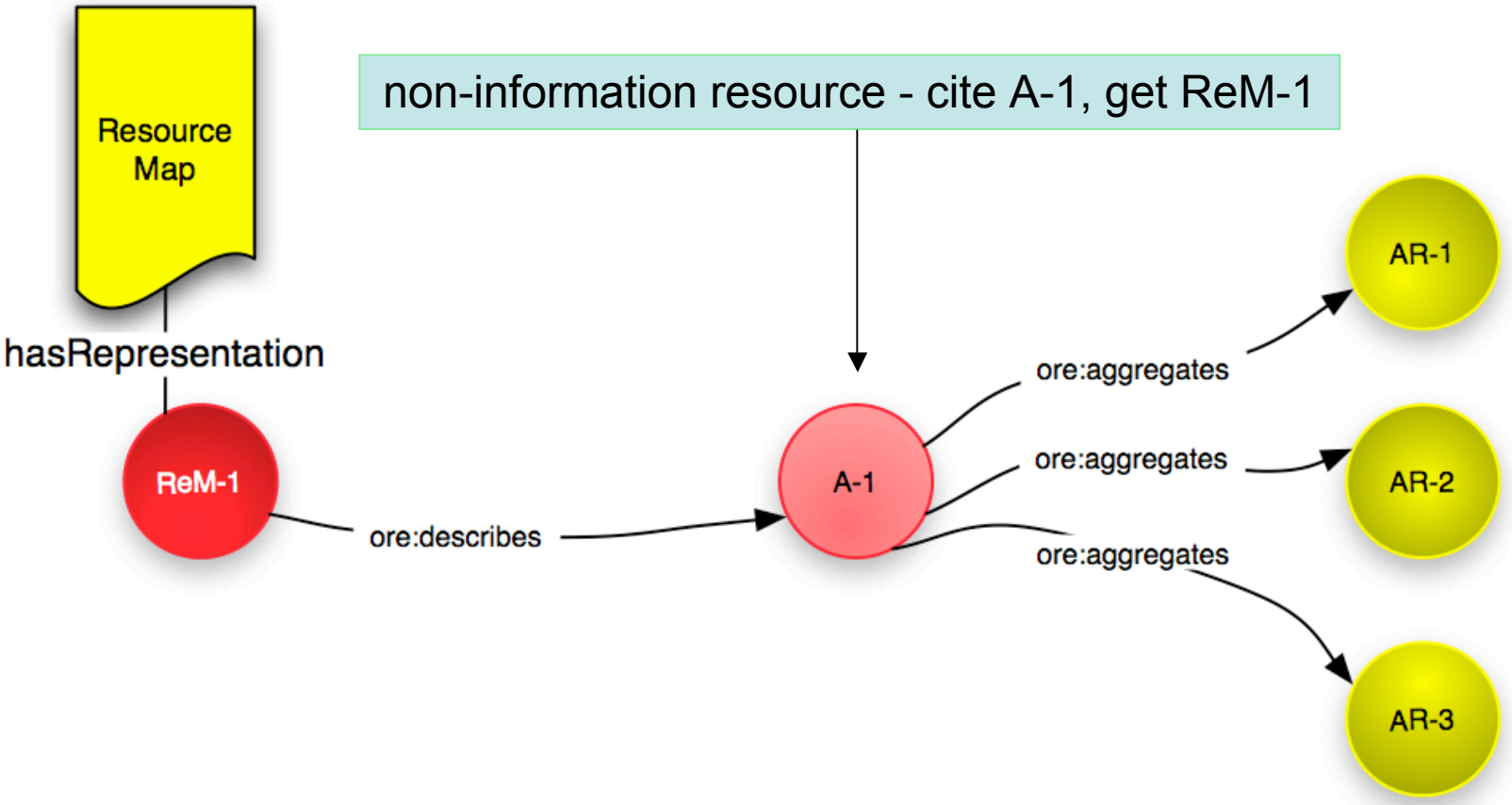
HTTP Implementation
Aggregation URI
Resource Map URI
Splash Page URI
Proxy URI

HTTP implementation



The Web is built of HTTP URIs

HTTP implementation



non-information resource - cite A-1, get ReM-1

Access to URI Aggregation yields a Resource Map: various approaches

Cool URIs

- Want simple, stable, manageable URIs
  - stability important for citation
- Certainly no technology baggage (.php, .asp etc.)
- Aggregation URI not tied to format of ReM

```
A-1   = http://example.org/foo
ReM-1 = http://example.org/foo.atom
```

Cool URIs with server support

Content Negotiation

- HTTP GET A-1
- Return ReM & ReM-1

```
Status: 200 OK
Content-Location: ReM-1

    <?xml version="1.0"...>
    ...ReM-1 in response...
```

303 Redirection

- HTTP GET A-1
- Redirect to ReM-1

(from Linked Data world)

```
Status: 303 See Other
Location: ReM-1
```

Cool URIs with server support: Multiple Resource Maps

- Choose one Resource Map as default (access from A-1 → authoritative)
- Indicate other ReMs with `ore:isDescribedBy` (chain of authority)

```
A-1   = http://example.org/foo
ReM-1 = http://example.org/foo.atom
ReM-2 = http://example.org/foo.rdf
```

Cool URIs without server support

- Use fragment identifier
- Also mentioned in CoolURI and Linked Data Specs
- Extension to multiple serializations is ugly

```
A-1   = http://example.org/foo.atom#aggregation
ReM-1 = http://example.org/foo.atom
```

RDFa

- “Splash Page” acts as human description and Resource Map
- “Splash Page” may be Resource Map and an Aggregated Resource (recommend different URIs)
- Resource Map URI MUST be different from Aggregation URI (as usual)

RDFa and the URIs (1)

- Cool URI

```
A-1   = http://example.org/foo
ReM-1 = http://example.org/foo.html   (+RDFa)
ReM-2 = http://example.org/foo.atom
```

- Fragment Identifier

```
A-1   = http://example.org/foo.html#aggregation
ReM-1 = http://example.org/foo.html   (+RDFa)
```

RDFa and the URIs (2)

- Fragment identifier to distinguish the Resource Map

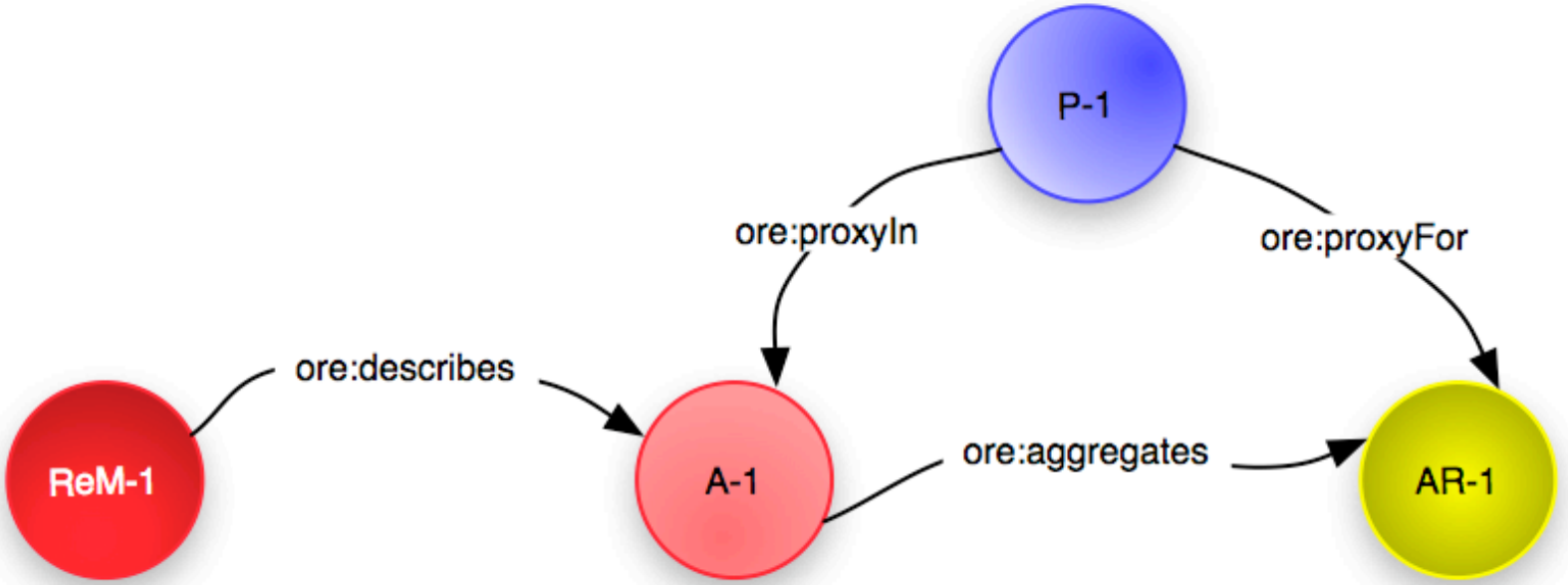
```
S-1   = http://example.org/foo.html       (+RFDa)
A-1   = http://example.org/foo
ReM-1 = http://example.org/foo.html#rem   (+RDFa)
ReM-2 = http://example.org/foo.atom
```

- 2 URIs for the same page

```
S-1   = http://example.org/splash.html    (== foo.html)
A-1   = http://example.org/foo
ReM-1 = http://example.org/foo.html       (== splash.html)
ReM-2 = http://example.org/foo.atom
```

Los Alamos NATIONAL LABORATORY

Proxy HTTP URIs



`ore:proxyFor` and `ore:proxyFor` to introduce a Proxy for an Aggregated Resource

Requirements for HTTP URIs for Proxies

1. Redirect to the Aggregated Resource with:
   HTTP status code "303 See Other" and Location: URI-AR
2. Indicate the Aggregation context with:
   HTTP Link header Link: <URI-A>; rel="aggregation"
3. No restriction on URI syntax, but...


ORE Proxy URI resolver

- Operated by OCLC (thanks!)
- Simple construction syntax:

  http://oreproxy.org/r?what=URI-AR&where=URI-A
  - parameter order important
  - careful to URI encode (potentially doubly)
- Resolver is compliant with required behaviour
- Allows the use of Proxy URIs at no extra cost

OAI Object Reuse and Exchange

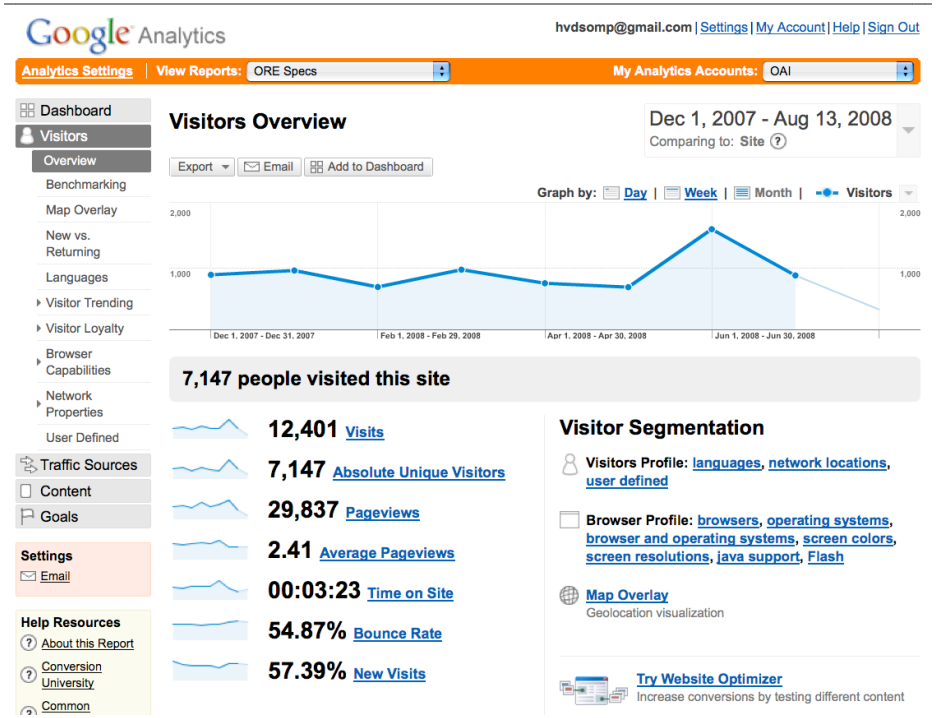
So what is going to happen with this all?

Kind of impossible to tell at this point.

But ...

What is going to happen with ORE?

- OAI-ORE sets a new direction to think about interoperability in our communities. We feel this is important as such.
  - Helps to align our community with the efforts towards a machine readable web: semantic web, linked data.



What is going to happen with ORE?

- Several experiments have explored applications based on ORE:
  - LANL:
    - digital preservation of aggregations
      - http://www.ctwatch.org/quarterly/multimedia/11/ORE_prototype-demo/
    - citation without citation manager
  - Mellon-funded:
    - UIUC: ORE for annotations
    - ODU/LANL: ReMember - social curation of aggregations
      - http://african.lanl.gov/preserve/
  - JISC-funded:
    - U Liverpool & HP: ORE for JSTOR content
    - U Cambridge: ORE for electronic chemistry thesis
  - Repository challenge at OR 2008
    - http://blip.tv/file/866653
  - RepoCamp ORE Challenge
    - http://docs.google.com/View?docid=dfsxqs8t_6hr85hfhh

What is going to happen with ORE?

- Several recent project proposals include ORE as a core infrastructural component:
  - NSF DataNet
  - Microsoft OREchem
  - …
- Significant interest in ORE from EU projects such as:
  - DRIVER
  - Europeana
  - EDLnet
- LANL's open source aDORe djatoka image server (JPEG2000) will do ORE …
  - **And, of course, you will all implement ORE!**

OAI Object Reuse and Exchange

So what is going to happen with this all?

Kind of impossible to tell at this point.

We should know more this time, next year …